Elucidating the allosteric mechanism of a mammalian glycolytic enzymeusing a computational and experimental biophysical approach

MacPherson, James Athol Stewart

Awarding institution:
King's College London

Elucidating the allosteric mechanism of a mammalian glycolytic enzyme using a computational and experimental biophysical approach.

James Athol Stuart Macpherson

A thesis presented for the degree of
Doctor of Philosophy in Biophysics

Randall Centre of Cell and Molecular Biophysics
School of Basic and Biomedical Sciences
King's College London
United Kingdom
September 2018

Abstract

Allostery is the regulation of protein function from a remote non-active binding site. Ligand binding incites a rapid process whereby information travels significant biomolecular distances to enact energetic changes to the the protein's functional centre. Despite the demonstrated importance of protein allostery in maintaining cellular homeostasis, investigating the mechanisms over the molecular temporal and spatial scales remains a challenge.

Fructose 1,6-bisphosphate (FBP) activates pyruvate kinase M2 (PKM2) through a well-described feed-forward process in glycolysis. Less understood, however, is how allosteric regulation of PKM2 occurs on a molecular level. Moreover, as several amino acids can compete for an additional binding site, it has thus far been unclear how concurrently bound ligands control PKM2 catalysis.

This thesis shows that, in a panel of cancer cell lines, the intracellular concentration of FBP exceeds that which is required to fully saturate binding to PKM2. Moreover, PKM2 is exposed to a dynamic concentration range of amino acids including the activator L-serine (Ser) and the inhibitor L-phenylalanine (Phe). When FBP is constitutively bound, Ser and Phe competitively regulate the maximal velocity of the PKM2-catalysed reaction, independent from changes to the protein's oligomeric state. To investigate the molecular determinants of multi-ligand regulation, a novel computational method *AlloHubMat* is developed and applied towards the analyses of molecular dynamics simulations of PKM2, identifying a number of 'allosteric hub' residues. A selection of the allosteric hub residues are mutated and an integrative approach is used to elucidate their impact on FBP- and Phe-mediated regulation of PKM2.

The findings herein demonstrate a role for residues involved in FBP-induced allostery in enabling the integration of allosteric input from Phe and reveal a mechanism that underlies the co-ordinate regulation of PKM2 activity by multiple allosteric ligands.

Acknowledgements

I would firstly like to thank my co-supervisors Prof. Franca Fraternali and Dr. Dimitrios Anastasiou for their supervision and in guiding the focus of the research. The work presented herein is the culmination of thousands of hours of dedicated work between the three of us. I thank them for their creativity in the pursuit of scientific thought.

I gratefully acknowledge the teaching and assistance of Dr. Laura Masino and Dr. Stephen Martin in experimental biophysical techniques and Dr. Jens Kleinjung in computational methods. I am also indebted to the members of my Thesis Committee: Dr. Paul Driscoll, Dr. Chris Lorenz and Dr. Markus Ralser. I thank them for volunteering their time and their thoughtful feedback throughout the progression of my project.

Native mass spectrometry and ion mobility measurements were performed together with Ms. Alina Theisen in the laboratory of Prof. Perdita Barran at the Manchester Institute of Biotechnology.

It was a fantastic pleasure to work along side Dr. Louise Fets, Dr. Patricia Figueiredo-Nunes, Dr. Fiona Grimm, Ms. Aakriti Jain and Ms. Natalie Bevan in the Anastasiou group; and Dr. Christian Margreitter, Ms. Anna Laddach, Ms. Irene Marzuoli, Mr. Josef Ng and Mr. Marius Kausas in the Fraternali group. I thank them all for their technical and conceptual suggestions, and for their friendship. I especially thank Dr. Louise Fets, with whom measurements of the cellular concentrations of PKM2 ligands were performed.

Finally, I thank my parents Prof. Andrew Macpherson and Dr. Anja Macpherson, and my brothers Mr. Atholl Macpherson and Mr. Alexander Macpherson, and Ms. Julia Kete for proofreading my manuscript and for their support.

This work was funded by a MPhil/PhD studentship between King's College London and the Francis Crick Institute, which receives its core funding from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust.

Contents

1 Introduction **12**

1.1 Enzyme catalysts are required to sustain cellular life . . . 12

1.1.1 The evolutionary emergence of biological catalysts . . . 13

1.1.2 Metabolic homeostasis is maintained by allosteric regulation . . . 16

1.2 Metabolite-autonomous allosteric regulation of pyruvate kinase . . . 19

1.2.1 Evidence for the involvement of PKM2 regulation in supporting pro-tumourigenic growth . . . 22

1.2.2 PKM2 protein structure . . . 27

1.2.3 Kinetics of ligand binding and regulation . . . 31

1.2.4 Concurrent regulation of PKM2 activity by multiple ligands . . . 32

1.2.5 Regulation of PKM2 structure and oligomerisation . . . 34

1.3 Investigating the role of protein dynamics in allosteric regulation . . . 38

1.3.1 Enzyme regulation can involve structural or vibrational motions . . . 38

1.3.2 Allostery is an ensemble phenomenon . . . 45

1.3.3 Identification of allosteric pathways from molecular dynamics simulations 48

1.4 Summary of research questions and the contribution of this thesis . . . 55

2 Theoretical, computational and experimental methods **59**

2.1 Theoretical principles of molecular dynamics simulations . . . 59

2.1.1 The Hamiltonian of a classical mechanical system . . . 60

2.1.2 The force calculation and integrating the equations of motion . . . 62

2.1.3 Molecular dynamics in various ensembles . . . 64

2.1.4 The functional form of the potential energy . . . 68

2.2 Parameters used for molecular dynamics simulations of PKM2 . . . 74

2.2.1 Molecular dynamics simulations in explicit solvent . . . 74

2.2.2 Molecular dynamics simulations *in vacuo* . . . 75

2.3 Protein biophysics . . . 76

2.3.1 Recombinant protein expression and purification . . . 76

2.3.2 Spectrophotometric assay to determine molar amounts of fructose 1,6-bisphosphate . . . 79

2.3.3 Measurement of PKM2 steady-state enzyme kinetics . . . 79

2.3.4 Analysis of the steady-state kinetics of PKM2 enzyme activity inhibition by Phe . . . 81
2.3.5 Measurement of the allosteric coupling co-efficient . . . 82
2.3.6 Circular dichroism spectroscopy . . . 83
2.3.7 Measurements of FBP binding to PKM2 . . . 84
2.3.8 Microscale thermophoresis measurements of phenylalanine and serine binding . . . 88
2.4 Native mass spectrometry . . . 89
2.4.1 Mass-deconvolution of native spectra . . . 89
2.4.2 Ion mobility mass spectrometry . . . 90
2.5 Measurement of metabolite concentrations in human cell lines . . . 91
2.5.1 Liquid chromatography-mass spectrometry (LC-MS) detection of metabolites . . . 91
2.5.2 Metabolite extraction and cell volume calculations . . . 92
2.6 Absolute quantitation of intracellular PKM2 amounts using targeted proteomics 94
2.6.1 Trypsin digestion . . . 95
2.6.2 Analysis of peptides by LC-tandem MS (LC-MS/MS . . . 95

3 Enzyme kinetics and ligand binding biophysical studies reveal a combined role for multiple allosteric ligands in the regulation of PKM2. 96
3.1 Introduction . . . 96
3.2 Characterising the kinetics of FBP binding to and activation of PKM2 . . . 98
3.2.1 FBP binds to PKM2 with nanomolar affinity . . . 98
3.2.2 Technical note: the protein concentration determines the error associated with the K_D of PKM2-FBP binding . . . 103
3.2.3 FBP binding leads to an increase in the affinity of PKM2 for its substrate 106
3.2.4 A phosphotyrosine peptide binds competitively with FBP to PKM2 . . 108
3.3 FBP binding to PKM2 induces inter-protomeric cooperativity . . . 111
3.3.1 Titrating PKM2(WT) with the catalytically-dead PKM2(R72A) results in a non-linear decay in enzyme activity . . . 111
3.3.2 A numerical framework for modelling oligomeric enzyme cooperativity . 115
3.3.3 FBP binding induced inter-subunit cooperativity under conditions of limiting substrate concentrations . . . 116
3.4 FBP and amino acids combinatorially regulate PKM2 activity . . . 119
3.4.1 Phe and Ser *per se* are K-type modulators of PKM2 catalytic activity . 119
3.4.2 The mechanism of phenylalanine inhibition of PKM2 catalysis is distinct depending on FBP binding. . . . 121
3.4.3 Ligand-induced changes in affinity cannot explain the changed phenylalanine-inhibition of PKM2 by FBP . . . 125
3.5 Simultaneous regulation of PKM2 by multiple ligands is relevant for enzyme regulation in cells . . . 128
3.5.1 Intracellular FBP binding to PKM2 is constitutive whereas Phe and Ser bind reversibly . . . 128
3.5.2 The addition of exogenous FBP does not change PKM2 activity in HCT116 cell lysates . . . 131

3.5.3 High milli-molar phospho-peptide concentrations are required to reversibly displace FBP binding . . . 134
3.6 Conclusion . . . 136

4 Native mass spectrometry reveals the structural basis of distinct ligand modes of action. 137
4.1 Introduction . . . 137
4.2 FBP binding regulates oligomerisation of PKM2 . . . 139
4.2.1 PKM2 forms a mixture of oligomeric states . . . 139
4.2.2 PKM2 dimers form stably about the A-A' interface . . . 141
4.2.3 FBP induces PKM2 tetramerisation . . . 145
4.2.4 $PKM2^{Apo*}$ tetramers are formed from a mixture of apo and holo protein 147
4.2.5 FBP binding induces subtle changes to the shape of tetrameric PKM2 . 149
4.3 Inhibitory amino acids have distinct effects on tetramerisation depending on whether FBP is bound . . . 150
4.3.1 Phe addition does not disrupt FBP-induced PKM2 tetramerisation . . . 150
4.3.2 Phe and FBP synergistically promote PKM2 tetramerisation . . . 152
4.3.3 Analytical evidence for simultaneous binding of FBP and Phe to PKM2 153
4.3.4 Phe partially reverses PKM2 to an 'apo-like' conformation . . . 155
4.4 Conclusion . . . 156

5 Molecular dynamics simulations predict hub residues involved in the allosteric transition of PKM2. 157
5.1 Introduction . . . 157
5.2 Molecular dynamics simulations of PKM2 capture ligand-induced conformational changes . . . 159
5.2.1 An integration time step of 2 fs appropriately conserves the energy in monomeric PKM2 simulations . . . 159
5.2.2 FBP binding causes PKM2 monomers to sample two distinct conformational states . . . 162
5.2.3 B-domain closure traps highly resident water molecules in the active site 167
5.2.4 B-domain dynamics is correlated with structural changes to catalytic residues in the active site . . . 168
5.2.5 Comparative dynamics between tetrameric and monomeric PKM2 . . . 171
5.2.6 The configurational entropy of PKM2 does not change upon allosteric ligand binding . . . 175
5.3 Identification of PKM2 allosteric hub residues using a novel software AlloHubMat 177
5.3.1 Explicit identification of sub-states from MD simulations using information theory . . . 180
5.3.2 PKM2 sub-states cluster according to the liganded state of the MD simulation . . . 183
5.3.3 A disperse network of hub residues are predicted to propagate FBP-induced activation of PKM2 . . . 185
5.3.4 Design of allosteric hub mutants (AlloHubMs) . . . 190
5.4 Conclusion . . . 192

6 Allosteric mutants reveal distinct ligand pathways that combinatorially regulate PKM2. 193
6.1 Introduction . . . 193
6.2 Biophysical characterisation of the AlloHubMuts . . . 195
6.2.1 AlloHubMuts have the same secondary structure content as PKM2(WT) 195
6.2.2 AlloHubMuts at either the A-A' or C-C' interfaces show distinct thermodynamics properties . . . 197
6.2.3 R489L reduces the binding affinity to FBP . . . 202
6.3 Characterisation of the functional response of AlloHubMuts to allosteric ligands 204
6.3.1 Several AlloHubMuts attenuate the allosteric coupling between the FBP pocket and the active site . . . 204
6.3.2 AlloHub residues A327 and C358 mediate multi-ligand allosteric coupling 205
6.3.3 Native spectra of AlloHubMut uncouples oligomerisation from allosteric activation . . . 209
6.4 Conclusion . . . 212

7 Discussion 213
7.1 PKM2 is concurrently regulated by multiple allosteric ligands in a range of cellular conditions . . . 215
7.2 FBP and Phe regulate PKM2 through a functional cross-talk . . . 220
7.3 AlloHubMat reveals residues that mediate the cross-talk between FBP- and Phe-induced allosteric regulation . . . 223
7.4 Conclusion . . . 228

List of Figures

1.1 Biological catalysts. . . . 15
1.2 Allosteric regulation of glucose metabolism. . . . 21
1.3 A cartoon representation of the three-dimensional structure of PKM2. . . . 28
1.4 A linear sequence diagram of PKM2. . . . 29
1.5 The energetic landscape of protein folding and native protein dynamics. . . . 40
1.6 Typical rate coefficients as a function of temperature. . . . 42
1.7 The M32K25 structural alphabet as a fragment-based representation of protein structure and dynamics. . . . 52
1.8 Time line of historical discoveries in the fields of protein dynamics and allostery, and PKM2 biology and biophysics. . . . 58

2.1 Examples of bonded interaction potential energies for arbitrarily selected atom pairs. . . . 71
2.2 The catalytic mechanism of the pyruvate kinase-catalyzed conversion of phosphoenolpyruvate and ADP to pyruvate and ATP. . . . 80
2.3 Visualisation of a ligand-receptor binding curve fitted using the *ligBind* package. 87

3.1 FBP binds to PKM2 with nano-molar affinity. . . . 102
3.2 Simulated protein-ligand binding data show that the measured binding affinity is over-estimated when high concentrations of protein are used in the experiment. 105
3.3 Addition of FBP results in an increase in the substrate affinity of PKM2 for phosphoenolpyruvate. . . . 107
3.4 Fluor-M2tide competes with FBP for binding to PKM2. . . . 110
3.5 A mixture of wild-type and catalytically-dead PKM2 has a non-linear decay in activity. . . . 114
3.6 FBP binding results in inter-subunit cooperativity as defined in a numerical model. . . . 118
3.7 FBP and amino acids have distinct combined roles in regulating PKM2 activity. 120
3.8 Phenylalanine inhibition follows one of two distinct mechanisms, depending on the presence of FBP. . . . 123
3.9 Amino acids and FBP do not reciprocally affect binding. . . . 127

3.10 Cellular concentrations of PKM2 allosteric effectors reveal a physiological role for multi-ligand regulation. . . . 130
3.11 PKM2 activity from HCT116 lysates is not further activated by addition of exogenous FBP. . . . 132
3.12 Intracellular concentrations of phosphoenolpyruvate are predicted to be constitutively limiting for PKM2 binding. . . . 133
3.13 Milli-molar concentrations of phosphorylated peptide is required to out-compete FBP binding in cells. . . . 135

4.1 The concentration-dependence of PKM2 oligomerisation. . . . 140
4.2 Ion mobility measurements reveal that PKM2 dimers are stable about the A-A' interface. . . . 144
4.3 FBP addition induces a dose-dependent tetramerisation of PKM2. . . . 146
4.4 The mass-deconvolved spectrum of $PKM2^{apo*}$ reveals a mixture of *apo* and *holo* species. . . . 148
4.5 FBP binding changes the global collision cross section of PKM2 tetramers. . . . 149
4.6 Phe binding does not disrupt FBP-induced PKM2 tetramerisation. . . . 151
4.7 FBP and Phe synergistically promote PKM2 tetramerisation. . . . 152
4.8 FBP and Phe can simultaneously bind to PKM2. . . . 154
4.9 Phe partially reverses FBP-induced conformational changes. . . . 155

5.1 Energy fluctuations calculated for different integration time- steps. . . . 160
5.2 Conformational plasticity and essential dynamics of monomeric PKM2. . . . 164
5.3 Conformational equilibrium of FBP-induced cap closure from a 500 ns MD simulation. . . . 166
5.4 Water residence time at the active site of mPKM2-FBP. . . . 167
5.5 Active site residue H78 adopts a distinct side-chain conformation depending on the liganded state of the protein. . . . 169
5.6 Charged-charged interactions at the A-A' interface prevent lateral B-domain closure in tetrameric PKM2. . . . 174
5.7 Allosteric ligands do not change the configurational entropy of PKM2. . . . 176
5.8 Schematic of the AlloHubMat and software flowchart. . . . 179
5.9 Conformational sub-states are identified from the geometric progression of the mutual information matrix. . . . 182
5.10 PKM2 correlated motions cluster according to the liganded state of the simulation. 184
5.11 The sub-state mutual information matrix is log-normally distributed. . . . 186
5.12 Hub residues are predicted to propagate energy transfer along allosteric pathways. 189
5.13 Design of Allosteric Hub Mutants (AlloHubMuts). . . . 191

6.1 An integrated computational and experimental strategy for identifying residues which transmit allosteric information between a binding pocket and the active site of a protein. . . . 194
6.2 Far-UV circular dichroism spectra of AlloHubMuts suggest no change in secondary structure content. . . . 196

6.3 Thermal unfolding spectra of the AlloHubMuts fitted to a two-phase unfolding model. . . . 200
6.4 The thermal stability of PKM2(K305Q) and PKM2(R489L) is increased upon FBP addition. . . . 201
6.5 The AlloHubMuts have a nano molar affinity for FBP with the exception of PKM2(R489L). . . . 203
6.6 The AlloHubMuts perturb either FBP-active site coupling or FBP-Phe coupling. 207
6.7 Native spectra of the AlloHubMuts reveal a complex relationship between allosteric activation and oligomerisation. . . . 211

7.1 Proposed model of PKM2 regulation by amino acids. . . . 219
7.2 The location and function of the AlloHubMuts. . . . 227

List of Tables

1.1 Structures of human PKM2 deposited on the Protein Data Bank. 30

2.1 Forward primer sequences for single-point mutants of PKM2 77
2.2 Reverse primer sequences for single-point mutants of PKM2 78
2.3 Liquid-chromatography mass spectrometry analysis parameters used for the detection of ^{12}C and ^{13}C metabolite standards. 93
2.4 Proteotypic peptide standards used for absolute quantification of PKM2 in cell lysates. 94

3.1 Dissociation constants for PKM2 self-association and hetero-association with PKM2(R72A), under various liganded conditions. 113
3.2 Steady-state Michaelis-Menten kinetic parameters for $PKM2^{apo*}$, and following the addition of FBP, Phe and Ser. 124

4.1 Theoretically-determined masses of PKM2 and FBP 148
4.2 Maximum entropy mass estimation of tetrameric PKM2 from nESI-MS m/z spectra under native-like and high cone-voltage ionisation conditions. 148

5.1 Single precision NVE simulation without thermostat or barostat coupling. . . . 160
5.2 Summary of MD simulations of monomeric and tetrameric PKM2. 161
5.3 Available functions contained within the R package AlloHubMat1.0 178

6.1 Equilibrium thermal unfolding kinetic parameters of the AlloHubMuts. 199
6.2 Steady-state dissociation constants of the AlloHubMuts for FBP. 202
6.3 Steady-state enzyme kinetic parameters of the AlloHubMuts. 208

Chapter 1

Introduction

1.1 Enzyme catalysts are required to sustain cellular life

Life on earth is dependent on the chemical synthesis of molecules, which constitute the building blocks of cellular organisms. Early proto-organisms had to adopt a way in which they could ensure a constant nutrient supply, without which life would be unlikely to persist on geological time scales. The selective pressure for auxotropy witnessed the emergence of essential chemical reactions that convert naturally available nutrients into various molecules to support specific functions. This included the provision of a wide variety of molecules suitable for the many complex anabolic and catabolic reactions essential for the support of organismic growth and division such as sugars, lipids, nucleotides and amino acids. The ability to perform chemical synthesis is achieved through pathways of sequential reactions collectively known as metabolism, and subsequently rendered organisms non-reliant on specific geochemical conditions enabling them to colonise a diversity array of ecological niches.

1.1.1 The evolutionary emergence of biological catalysts

Metabolism can be viewed as a network of inter-connected chemical reactions, whose topological features are largely conserved in all studied organisms. Two important inferences can be drawn from the highly conserved nature of metabolic reactions: (i) metabolic networks likely reached a state of efficiency very early on in evolutionary time scales that could barely be improved further by natural selection, and (ii) all present forms of metabolism likely descend from the same original network[1]. These conclusions give rise to the hypothesis, originally proposed by Aleksandr Oparin in his 1938 book 'The Origin of Life'[2], that a common metabolic network topology dates back to non-enzymatic chemistry in the presence of simple inorganic ions as found in the highly reducing geochemical environment that made up the Archaen sediment. Indeed, metabolic intermediates of two pathways in central carbon metabolism - glycolysis and the pentose phosphate pathway (PPP) - were shown to undergo non-enzymatic interconversion in chemical compositions replicating the pre-biotoic Archaen ocean[3]. Additionally, ions that were abundant in the Archaean sediment, such as Fe^{2+} and PO^{3-}, can catalyse reactions which are reminiscent of the modern PPP, suggesting that modern metabolic reactions did not necessarily result from the evolutionary selection of complex enzyme catalysts, but rather that such reactions were able to proceed under prevailing chemical conditions in the Archaean sea[4].

Small molecules such as metal ions can efficiently catalyse many chemical reactions. Several constraints on biological systems, however, render their utility insufficient for supporting complex cellular life *per se*. Many metal ion catalysts found at high concentrations in the Archaean ocean, do not accumulate to sufficient concentrations in modern biological conditions, such that they can function as efficient catalysts. For example, ferrous iron is readily water soluble in the absence of oxygen, and anoxic conditions in the earth's early surface environment resulted in the Archaean milieu preserving a high iron content, whereas its oxidised form was readily depleted upon the great oxygenation event[5]. This likely imposed a strong selection pressure for species to evolve their metabolism, becoming less dependent on the presence of

increasingly rare and/or insoluble molecules, eventually leading to the emergence of macromolecular catalysts.

The RNA world hypothesis provides a conceptual framework whereby the emergence of genetic selection can be explained by RNA molecules acting as self-replicators, thereby unifying heritability and the propensity for stochastic evolution[6]. Indeed, RNA molecules have been used in directed evolution experiments and, in a milestone paper, Lincoln and Joyce (2009)[7] presented an experimentally designed *in vitro* self-replicating RNA enzyme. Nevertheless, the ability for RNA enzymes to catalyse reactions, other than phosphoester cleavage or ligation, is limited. The design of an RNA enzyme by Fusz *et al.* (2005)[8] that could act as an intermolecular catalyst of the aldol reaction was found to be Zn^{2+}-dependent. Peptides can also bind to metal ions, and simple metal-binding RNA or peptide structures could provide the starting point for positive selection by supporting multiple reactions, and would have prevented the loss of low concentrations of metal ions through diffusion and/or precipitation[4]. A likely consequence of non-enzymatic metabolism was that early enzyme and RNA catalysts were able to evolve in the background of non-specific metal ion-catalysed reactions. To avoid the production of superoxide from hydrogen peroxide by ferrous iron, modern cells evolved a complex iron transport mechanism to prevent accumulation of ferrous iron in the cytoplasm[9]. Subsequently, peptide and RNA secondary structures would have increased the affinity for some substrates, facilitating access to the retained metal catalyst, giving rise to early enzymes.

A dynamic chemical system operating at typical biological rates, with all intermediates at concentrations in the nanomolar to millimolar range, can be constructed only by use of efficient catalysts. The emergence of enzyme catalysts enabled metabolism to cope with the high demand for biomass, necessary to sustain the requirements of cellular organisms (**Fig. 1.1**). Enzymes can achieve unrivalled substrate specificity due to their complex protein folds and can distinguish between very similar compounds, which prevents the progression of undesirable side-reactions[4]. In addition to greater fidelity, the phenomenon of enzyme substrate

specificity frequently results in the binding of metabolites, which closely resemble the natural substrate, to the active site of the enzyme, resulting in a steric preclusion of the natural substrate from the catalytic site. This phenomenon manifests as competitive inhibition, and provides an example of how the kinetic properties of enzyme catalysts can be reversibly regulated. Compared to small-molecule catalysts, whose catalytic activity is not readily regulated, the complex three-dimensional structure of enzymes houses pockets which allow metabolites and other small molecules to bind and regulate enzyme activity. The implications of enzyme regulation for metabolism are discussed in the following section.

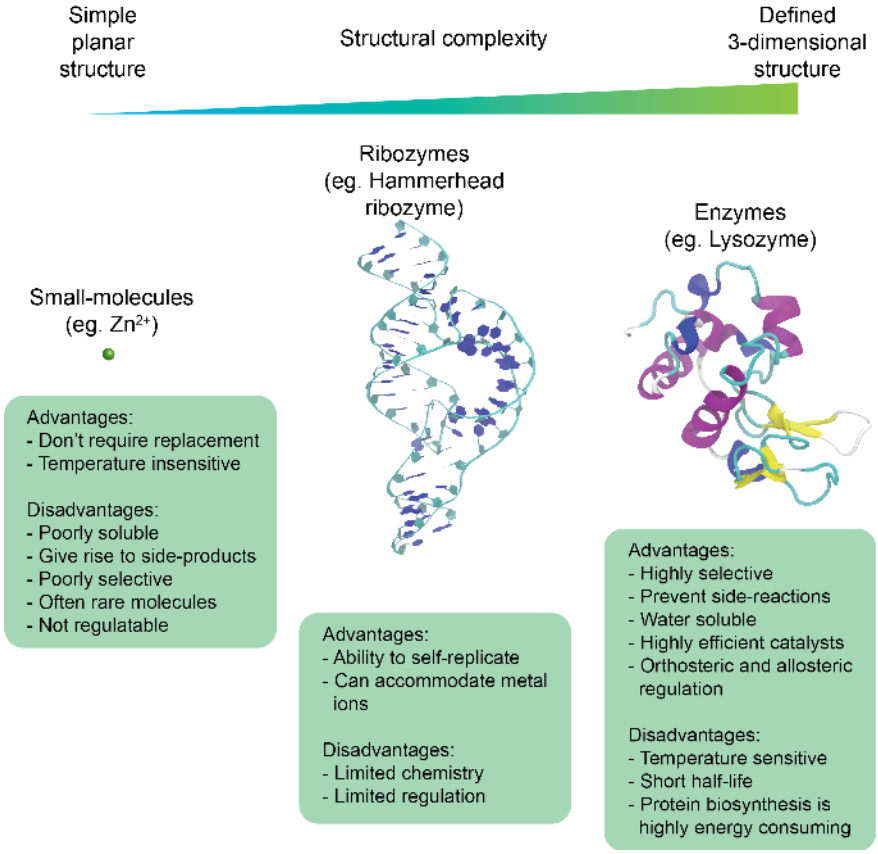

Figure 1.1: ***Biological catalysts.*** *An illustration of the structural complexity of small molecule catalysts, ribozymes and enzymes, and their advantages and limitations.*

1.1.2 Metabolic homeostasis is maintained by allosteric regulation

The *in vivo* capacity of an enzyme to realise a particular catalytic flux, at steady-state, is determined by its abundance and kinetic properties. Enzyme abundance can be regulated by gene expression, whereas the kinetic properties of enzyme-catalysed reactions are regulated post-translationally through covalent modifications under the control of signalling pathways and via interactions of enzymes with small molecules. The average concentration of a metabolite in central carbon metabolism is of the order of 1 mM, while glycolytic flux varies between 0.1 and 0.3 (mM· s^{-1}).[10] Assuming steady-state conditions, where the rate of metabolite production is approximately equal to the rate of consumption, the rate of turnover of the pool of metabolites (r) can be determined:

$$r = \frac{P}{v} \qquad , where \quad \frac{\partial P}{\partial t} = 0 \tag{1.1}$$

where P is the pool size, v is flux and t is time. The turnover time of metabolites in glycolysis therefore has an approximate range in mammalian cells between 2.8 s and 0.93 s. A quantitative model of time-dependent gene expression changes in mouse fibroblast cells estimated a median transcription rate of approximately 3 mRNAs per hour and and a median translation rate of about 140 proteins per mRNA per hour[11]. Therefore, if the reactions of metabolite consumption and production are not in balance, the metabolite pool will become rapidly consumed (or compounded) before gene expression can affect changes to enzyme activity by changing expression levels. Uncontrollable flux imbalances in cells would precipitate an unsustainable biological situation where the metabolite pool would be converted into a handful of compounds, consistent with the end-products of the most energetically favoured enzyme-catalysed reactions in the metabolic network. In essence: metabolism cannot afford to wait for gene expression changes. Instead, the kinetic properties of metabolic enzymes must be tightly regulated to ensure homeostasis.

Feedback inhibition in metabolism was first hypothesised by Novick and Szilard (1950)[12],

when studying the regulation the tryptophan synthesis pathway, and has since been found in almost every biosynthetic pathway. Computational models of simple metabolic networks found that product-feedback inhibition enables optimal growth of cells by minimizing futile cycling and optimizing fluxes[13]. Goyal *et al.* (2010)[13] showed that, while feedback inhibition is sufficient to control fluxes, the effectiveness of simple product-feedback inhibition comes at the cost of producing high levels of certain metabolite pools, which would likely cause toxicity and osmotic imbalance in cells. Large metabolite pool sizes can be restricted if feedback inhibition is ultra-sensitive. In addition to active (orthosteric) sites, complex enzyme structures present additional *allosteric* binding pockets to which metabolites and other small molecules can bind. Subsequent to binding to an allosteric pocket, information can be propagated over significant biomolecular distances to enact changes to the active site, thereby changing the kinetic properties of the enzyme. This mechanism is known as allosteric regulation and has the potential to both inhibit and activate enzyme activity on very short temporal scales (typically in the range of ns to ms), thus providing a highly sensitive regulatory mechanism for the maintenance of homeostasis.

Tight regulation of enzyme catalysis on short timescales is particularly evident in glycolysis, where feed-forward and feed-back regulation manifests in rapidly oscillating concentrations of glycolytic metabolites. This oscillatory behavior of glycolysis was first discovered in yeast cells[14]. In early investigations, glucose was fed at a continuous rate to yeast cell extracts and oscillations were observed in the concentrations of glycolytic intermediates with a frequency of several minutes[15]. Acute changes to the composition of the extracellular environment also leads to rapid changes in mammalian cell glycolysis, which can be rationalised by enzyme regulation on short temporal scales. Empirical evidence for this hypothesis has been provided recently by studies using fluorescence sensors to monitor the redox balance between nicotinamide adenine dinucleotide in its oxidised (NAD^+) and its reduced (NADH) forms, as representative of the balance between oxidative and reductive metabolism. Using a genetically-encoded fluorescence peredox sensor, Hung *et al.* (2011)[16] found that the exchange between cytosolic NADH and

NAD^+ occurred within minutes of acute changes to culture media of mouse neuroblastoma cells, suggesting that glycolytic flux rapidly reaches the steady-state. Taken together, the rapid equilibration of cellular metabolites in response to changed extracellular conditions shows that metabolic homeostasis is maintained on far shorter time scales than can be achieved by gene expression changes, requiring allosteric regulation of enzyme catalysts.

Due to the importance of allosteric regulation in maintaining homeostatic control of cellular metabolism, subtle changes to the regulation of metabolic enzymes can result in significant alterations to the phenotype of a cell. This is well exemplified in several disease states including cancer, where characteristic changes to central carbon metabolism have been shown to play essential roles in sustaining rapid growth and proliferation. Some pro-tumorigenic features of metabolism can be attributed to changes in gene expression, such as oncogenic mutations in the Ras/Erk pathway resulting in enhanced glucose uptake. Ras-driven over expression of hexokinase-2 in 3T3 cells was shown to lead to increased rates of glycolytic flux, thus increasing the rate of turnover of the metabolite pool by up to 30 %[10]. This introduces a greater requirement of cancer cells for allosteric regulation of metabolic enzymes, to maintain metabolic homeostasis in the context of increased flux. Indeed, many of the changes to metabolic reprogramming in cancer cells have been linked to altered allosteric regulation of key metabolic enzymes. Changes to enzyme regulation is further compounded by aberrant growth factor signalling, which can alter enzyme activity through post-translational covalent modifications, and by changes to nutrient availability in the tumour microenvironment, which can result in changes to intracellular metabolite pools.

1.2 Metabolite-autonomous allosteric regulation of pyruvate kinase

Mammalian glycolysis involves the conversion of the six-carbon sugar glucose into the three-carbon keto-acid pyruvate, through a series of ten enzyme-catalysed reactions (**Fig. 1.2**). In addition to providing cellular energy in the form of adenosine triphosphate (ATP) and pyruvate, which can be further oxidised to form acetyl-CoA that enters into the TCA-cycle, glycolytic intermediates also provide precursors for biomass production. Regulation of glycolytic flux plays an important mechanistic role in mammalian physiology, contributing not only to circulating glucose homeostasis and providing ATP, but also to the provision of biomass building blocks in contexts such as cell proliferation, immune activation and angiogenesis. Glucose-utilising pathways are, in turn, coupled to the interconversion of several co-factors including NAD(P)H that contribute to reduction-oxidation homeostasis.

Regulation of glycolysis is principally achieved through allosteric regulation of several allosteric enzymes including hexokinase (HK), phosphofructokinase (PFK-1) and pyruvate kinase (PK). Pyruvate kinase exists as one of four mammalian isoforms (PKM1, PKM2, PKL and PKR), all of which catalyse the interconversion of PEP and pyruvate in the terminal step of glycolysis. Differential splicing produces L- and R-type PK mRNA from the *pkl* gene. The expression of PKL is limited to liver and some cells of the pancreas, intestine and kidney, and PKR is expressed in erythrocytes. The human pkm gene has 12 exons with either the 9^{th} or the 10^{th} alternatively spliced to generate the PKM1 or PKM2 transcripts, respectively. Both the PKM1 and PKM2 transcripts are identical in length but encode a 56-amino acid region that differs at 22 residues. This localised difference in the primary sequence forms an allosteric pocket in the C-domain of PKM2 to which an activator fructose 1,6-bisphosphate (FBP) can bind.

In contrast to the M1 isoform, which is thought to be constitutively active in cells[17], the

catalytic activity of PKM2 can be modulated by several endogenous metabolites that bind to one of two identified allosteric pockets. Allosteric activators include the up-stream glycolytic intermediate FBP[18]; two amino acids L-serine (Ser)[19] and L-histidine (His)[20]; and SAICAR[21], an intermediate in purine synthesis. Conversely, the enzyme activity of PKM2 can be inhibited L-tryptophan (Trp)[20], L-alanine (Ala)[20], L-phenylalanine (Phe)[22], a phosphotyrosine peptide[23] and T_3 hormone[24]. Expression of the M2 isoform is dominant in many adult tissues, and is also expressed in healthy proliferative tissues such as embryonic tissue, lymphocytes and intestinal epithelial cells[25]. Additionally, PKM2 is highly expressed in tumour cells[26], and the observation that elevated expression of PKM2 correlates with a poor clinical prognosis[27,28] precipitated efforts to inhibit the activity of PKM2 in an attempt to reduce the elevated glycolytic flux seen in many tumours[29,30]. A landmark study by Christofk *et al.* (2008)[26], however, revealed that PKM2 over-expression could promote metabolic adaptations to nutritional stress in cell culture and increase tumour growth. This led to the prevailing view that lower PK activity, through the enhanced expression and regulation of the M2 isoform, is beneficial for anabolic metabolism and thus for cell proliferation and tumour growth[25,31–35].

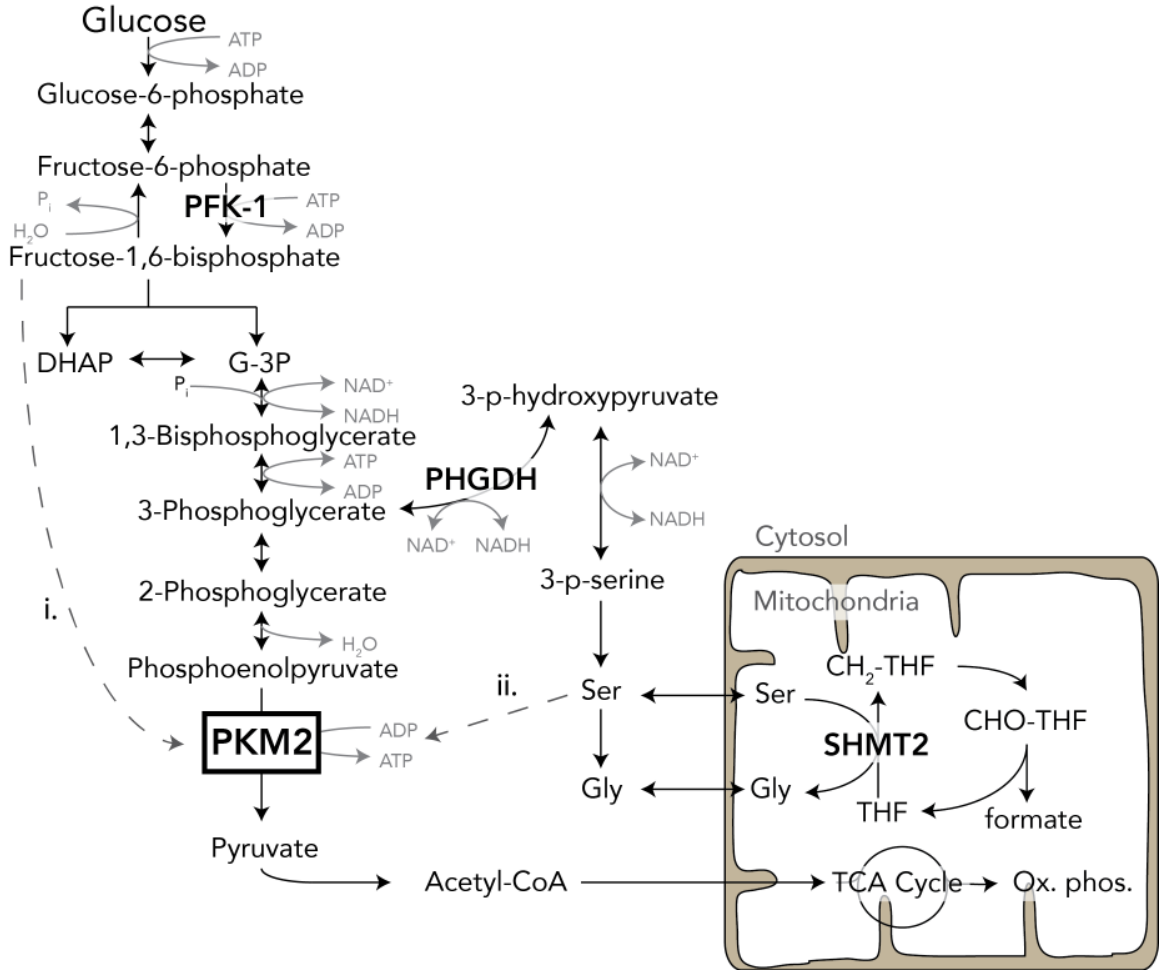

Figure 1.2: ***Allosteric regulation of glucose metabolism.*** *PKM2 is allosterically regulated by D-fructose-1,6-bisphosphate (FBP; an up-stream glycolytic intermediate) and by L-serine (Ser). Increased glucose uptake results in an accumulation of up-stream glycolytic intermediates. From the current paradigm it follows that accumulation of FBP leads to enhanced allosteric activation of PKM2, through a feed-forward mechanism [i]. Decreased serine uptake due to nutrient deprivation results in reduced binding of serine to PKM2, and hence reduced allosteric activation. This spares glucose carbons for de novo serine synthesis and one-carbon metabolism.*

1.2.1 Evidence for the involvement of PKM2 regulation in supporting pro-tumourigenic growth

A number of studies have revealed several mechanisms by which allosteric regulation of PKM2 activity contributes to tumour physiology. Growth-factor signalling pathways have been shown to play a major role in programming metabolic pathways in cells by mediating acute as well as long-term changes in cell metabolism. Activation of protein tyrosine-kinases had previously been shown to result in a decrease in the enzyme activity of PKM2. Christofk *et al.* (2008)[23] elucidated the mechanism by which protein-tyrosine kinases acutely regulate the activity of PKM2, finding that aberrant phosphorylation of proteins at tyrosine residues inhibits PKM2 enzyme activity by precipitating the release of the allosteric activator FBP[23]. To examine the effect of phosphotyrosine protein binding on FBP-bound PKM2, the authours obtained a peptide binding motif for PKM2 by screening a phosphotyrosine-enriched peptide library matrix to identify novel phosphotyrosine (pTyr)-binding proteins from cell lysates. Christofk *et al.* then synthesized both the phosphorylated and unphosphorylated versions of the optimal peptide: P-M2tide (GGAVDDDpYAQFANGG) and NP-M2tide (GGAVDDDYAQFANGG), respectively for *in vitro* experimentation resembling the domain of a protein tyrosine-kinase with the capacity to bind to PKM2. Moreover, PKM2 inhibition via a signalling-induced protein phosphotyrosine-binding event led to the diversion of glucose carbons away from energy-producing pathways and into lactate and lipid production[23].

Additional environmental conditions associated with the tumour microenvironment such as hypoxia, matrix detachment and inflammation can all lead to excess production of reactive oxygen species (ROS)[36–38], which at sufficiently high concentrations can damage cellular components and compromise cell viability[39]. In this context, regulation of PKM2 has been shown to bestow cells with pro-tumourigenic survival advantages. Oxidation of PKM2 at C358 inhibits enzyme activity, promoting the utilisation of glucose carbons through the pentose phosphate pathway to sustain sufficient reducing potential for the detoxification of ROS[40].

Allosteric inhibition of PKM2 activity can facilitate the utilisation of glucose carbons for *de novo* serine synthesis and one-carbon metabolism. Though generally classified as a nutritionally dispensable amino acid, serine plays an essential role in a number of metabolic processes including the synthesis of other amino acids, lipogenesis and methylation reactions that occur via the generation of S-adenosylmethionine[41]. Furthermore, the folate cycle, a serine-catabolising process, can regenerate essential cofactors such as NAD(P)H and ATP. As tumour growth supersedes its vascular support, uptake of serine becomes necessary for cancer cell survival. The transfer of human colon cancer cells from culture conditions containing serine to serine-deprived media was found to suppress aerobic glycolysis and enhance the utilisation of glucose carbons in energy producing metabolic pathways[42]. The preference for high-flux aerobic glycolysis in the context of enhanced serine uptake is, in part, driven by allosteric regulation of PKM2. Serine is an allosteric activator of PKM2[19], and reduced activation of PKM2 enables cancer cells to adapt to life under limiting nutrient conditions by allowing for the accumulation of glycolytic intermediates, which are subsequently diverted into serine metabolism through phosphoglycerate dehydrogenase (PHGDH). This mechanism ensures that rapidly proliferating cells to sense and respond to availability of serine, thereby retaining a reserve pool of an important biosynthetic precursor.

In addition to fluctuations in extracellular concentrations of serine due to limited nutrient supply in poorly vascularised tumors, serine availability can also be limiting due to enhanced activity of enzymes involved in one-carbon metabolism. Rapidly proliferating cancer cells demonstrate an enhanced requirement for nucleotides, resulting from DNA replication stress[43]. Mitochondrial serine hydroxymethyltransferase 2 (SHMT2) catalyses the conversion of serine to glycine, while simultaneously hydrolysing tetrahydrofolate (THF) into 5,10-methylenetetrahydrofolate, a co-factor used in purine biosynthesis (**Fig. 1.2**). An investigation into the role of serine metabolism within the ischaemic zones of gliomas found that high SHMT2 facilitates a cross-talk between nucleotide biosynthesis and glycolysis by sequestering serine, FBP and SAICAR; thereby reducing oxygen consumption and eliciting a

metabolic state that confers a survival advantage to cells in hypoxic tumor regions[44]. Further in support of this hypothesis, deletion of PKM2 in non-immortalized primary cells from PKM2-conditional mice[45] lead to impaired DNA synthesis and cell cycle progression, resulting from severe thymidine depletion[46]. Limiting nucleotide biosynthesis in PKM2-null cells could be rescued by supplementation with exogenous thymidine, suggesting that expression of PKM2, and regulation of its enzyme activity, is required to support DNA synthesis in the context of exponential cell growth[46].

The mechanistic link between low pyruvate kinase activity and cancer cell proliferation promoted drug-discovery efforts, which identified several small-molecule activators that selectively target PKM2[47–49]. Exogenous PKM2 activators were found to bind to an allosteric pocket, distinct from other binding pockets for FBP and amino acids, and resulted in an increased affinity for the substrate PEP[17]. An investigation into potential *in vivo* therapeutic effects of the PKM2 activators found that the treatment of mice baring H1299 tumour xenografts with one of the activators Tepp-46, resulted in an inhibition of the xenograft tumors relative to the vehicle-treated control animals[17]. Impaired tumour growth was accompanied by decreased levels of serine, lactate and ribose phosphate (an intermediate of the pentose phosphate pathway) and increased acetyl-CoA, suggesting that activation of PKM2 resulted in metabolic changes that were incompatible with pro-tumourigenic growth[17,50].

Taken together, these observations demonstrate that expression of PKM2 affords cancer cells the increased metabolic plasticity to respond to varying environmental conditions that a constitutively high or low activity PK enzyme would not readily allow. Nevertheless, the differential requirements of cancer cell metabolism, and how this is supported by pyruvate kinase expression, remain under investigation. Israelsen *et al.* (2013)[45] generated mice with a conditional allele that abolished PKM2 expression without disrupting PKM1 expression, which were then crossed with a *Brca1*-loss-driven breast cancer mouse model. Intriguingly, mice *without* expression of PKM2 showed an accelerated tumour-associated mortality compared to the wild-type

control, suggesting that expression of PKM2 is not required for cancer cell growth and indeed its loss may promote tumour progression under certain conditions. Germline PKM2-*null* mice were found to spontaneously develop multiple macroscopic liver tumours[51]. More recently, Morita *et al.* (2018)[52] reported novel engineered mouse models that express either PKM1 or PKM2, at protein levels equivalent those found endogenously in tissues. The authors found that, across numerous cancer models, expression of PKM1 promoted and PKM2 inihibited tumourigenesis relative to the wild-type control[52].

The studies by Israelsen *et al.* (2013)[45], Dayton *et al.* (2016)[51] and Morita *et al.* (2018)[52] highlight a potentially complex relationship between PK expression and cell proliferation. The use of genetic tools to study the role of a metabolic gene in a complex disease are, however, limited in by their binary nature (i.e. comparing a phenotype produced by the knock-out of a single gene with that of wild-type gene expression). In contrast, metabolism, a product of the complex interaction between many genes and the nutritional environment, constitutes a quantitative continuum of phenotypes. Furthermore, spontaneous development of hepatocellular carcinoma in $PKM2^{-/-}$ models is characterised by the depletion of the nucleotides guanosine and cytidine[51], which may suggest impaired DNA replication resulting from decreased glucose utilisation in nucleotide biosynthesis. This, in addition to high systemic inflammation and metabolic changes reported by Dayton *et al.* (2016)[51], may explain why the mice develop tumours in old age.

While the metabolic conditions under which PKM2 exerts pro-tumourigenic effects are controversial, allosteric regulation of PKM2 activity remains an important feature of glycolytic control. Crucially, expression of the M2 isoform enables cells to undergo dynamic metabolic changes, which support pro-tumourigenic growth in a number of contexts. Nevertheless, the findings presented by Morita *et al.* (2018)[52] suggest that the metabolic outcome of cells may not necessarily map to a single enzyme activity status, but rather that both enhanced and decreased pyruvate kinase activity can promote or suppress tumorigenesis, depending on the

environment and status of all the other genes in the metabolic network[53]. Therefore, an understanding of the molecular basis of PKM2 regulation is critical, though this remains largely elusive both in terms of how conformational changes are induced following the binding of allosteric ligands and how PKM2 is able to integrate activating and inhibitory signals after binding to a litany of endogenous allosteric effectors. In this context, the structural and biophysical aspects of PKM2 regulation are reviewed in the following section.

1.2.2 PKM2 protein structure

Mammalian PKM2 is a homo-tetrameric protein, each protomer containing four structural domains: the A-, B-, C- and N- domains (**Fig. 1.3**). The protomers within the tetrameric assembly of PKM2 are associated with 222 symmetry. The 222 symmetry is exact only for the A- and C-domains, whereas the B-domains of subunits 1 and 2 are not related to each other by exact two-fold axes because of their different orientations. The A-domain is the largest of the four, spanning from residues N44 to G116 and P219 to I389, and forms a symmetric $\alpha_8\beta_8$ TIM barrel. The active site is located within the A-domain, underneath the B-domain cap (**Fig. 1.3**). The B-domain, consisting of a stretch of residues between P117 and L218 forming eight antiparallel β-sheets, is mobile and closes on the active site upon binding of the Mg^{2+}-ADP substrate complex[54]. The C-domain is composed of residues 390-531 and forms five α helices and five β sheets. A linear sequence diagram showing the positions of the domains is shown in **Fig. 1.4**. The two-fold symmetry of the homo-tetramer gives rise to two unique protomeric interfaces: the A-A' and the C-C' interfaces. The A-A' interface is formed from the A-domains of two adjacent protomers, consisting of a total of 75 residues with a total surface area of 2.7 nm^2. There are an estimated 34 hydrogen bonds and 14 salt-bridges that span across the A-A' interface forming contacts between the adjacent protomers. The C-C' interface is less than half the size of the A-A' interface, and is formed of the C-domains of two adjacent protomers in the tetramer assembly, with a total surface area of 1.1 nm^2. An estimated 17 hydrogen bonds and 11 salt-bridges are formed across the C-C' interface. FBP binds to a pocket within the C-domain of PKM2 (**Fig. 1.3**), proximal to the C-C' interface and forms interactions with W482, R489, K433, T432, S434, S437, G520 and F521[55]. The N-domain is composed of the first 46 residues in the protein that form a helix-loop-helix, connected to the A-domain by a flexible linker between L33 and G46. The N-domain forms an apolar allosteric pocket to which the small molecule synthetic activators bind, forming contacts with F26, Y390 and L394[17].

A number of amino acids competitively binding to an allosteric pocket, which is sandwiched

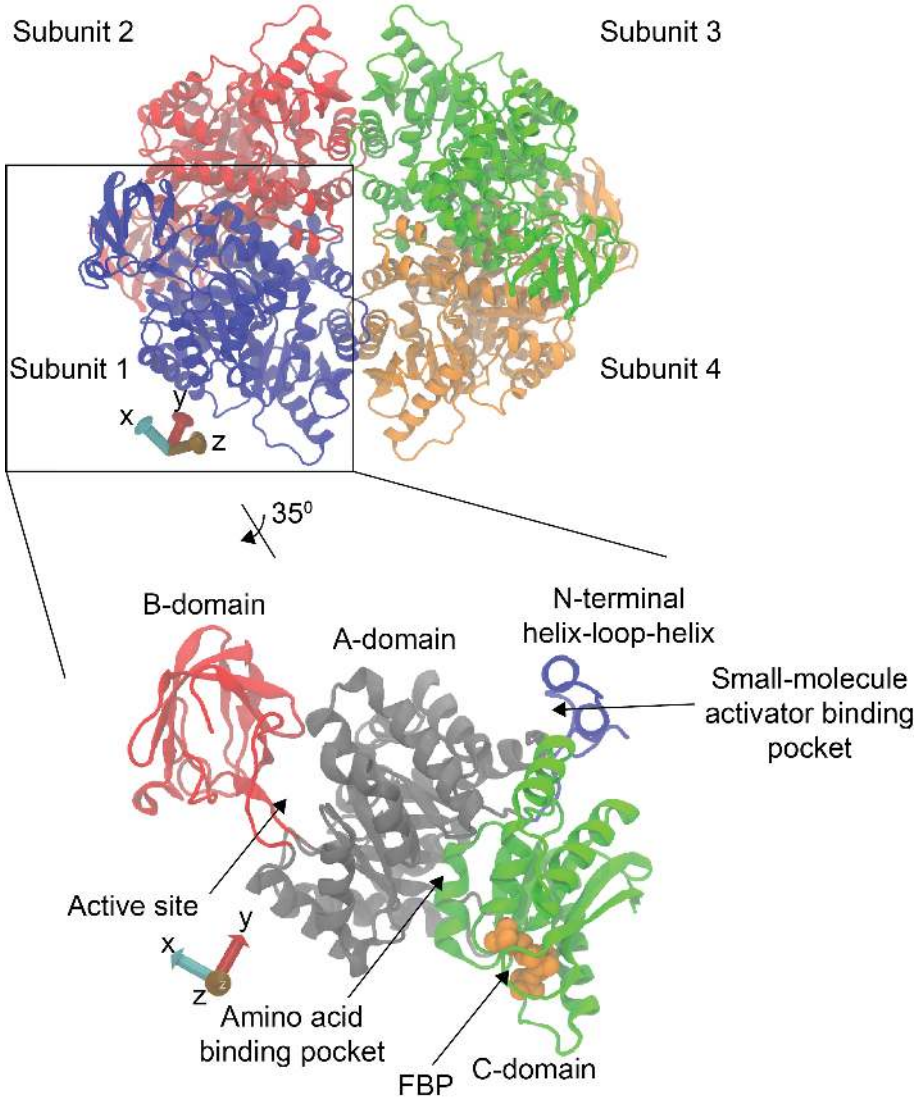

Figure 1.3: ***A cartoon representation of the three-dimensional structure of PKM2.*** *The tetrameric structure of PKM2 is shown above with subunits 1, 2, 3 and 4 of the homotetramer coloured in blue, red, orange and green, respectively. The structure of one of the four monomers within the tetrameric PKM2 complex is shown below. Each PKM2 monomer consists of four structural domains, which are coloured in the cartoon structure. The N-terminal helix-loop-helix is shown in blue, the A-domain is shown in grey, the B-domain is shown in red and the C-domain is shown in green. The C-domain contains the binding pocket for fructose 1,6-bisphosphate, which is shown in an orange space-fill representation. Locations of the active site, the amino acid binding pocket and the small-molecule activator binding pocket are indicated.*

between the A- and C-domains[20] (**Fig. 1.3**). Co-crystal structures of PKM2 in complex with either L-histidine (His), L-serine (Ser), L-alanine (Ala), L-tryptophan (Trp) or L-phenylalanine

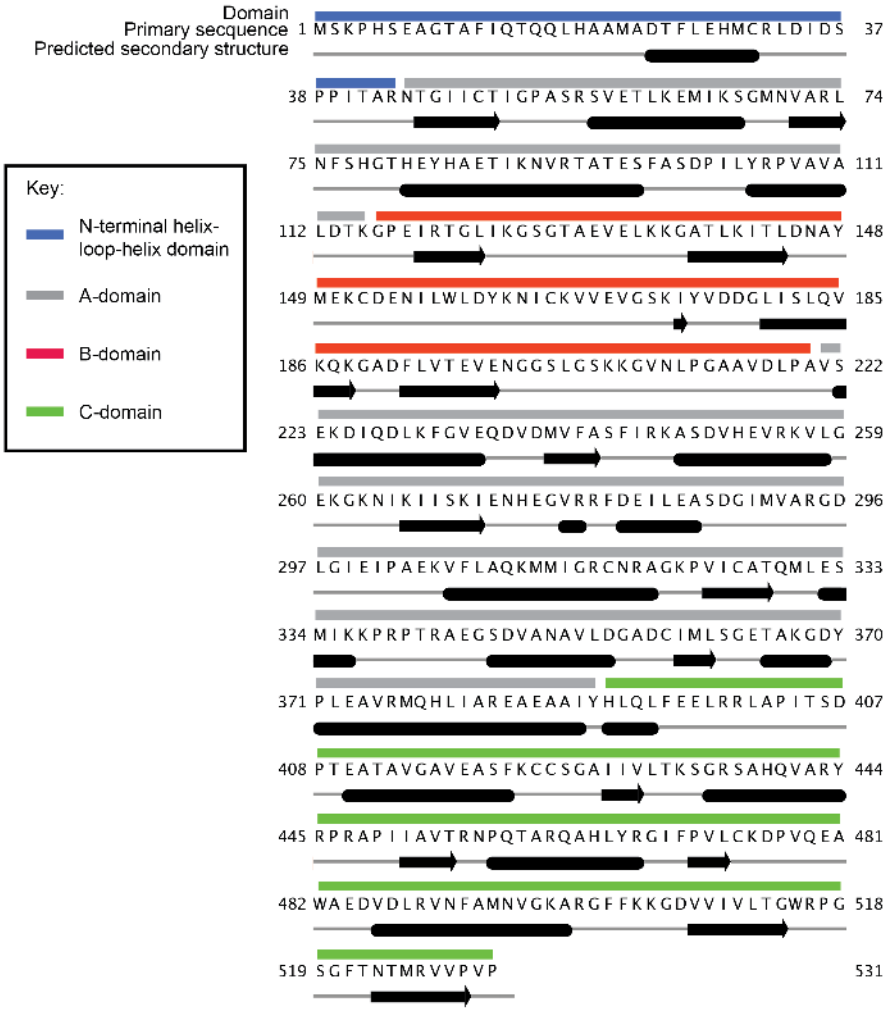

Figure 1.4: ***A linear sequence diagram of PKM2.*** *The full primary sequence of human PKM2 is annotated with the domain composition and the predicted secondary structure of the protein. The domain composition of the primary sequence is indicated by coloured bars above the protein sequence; the N-terminal helix-loop-helix is indicated with a blue bar, the A-domain with a grey bar, the B-domain with a red bar and the C-domain with a green bar. The predicted secondary structure is shown below the primary sequence; helices are indicated with cylinders and sheets are indicated with thick arrows.*

(Phe), found similar binding poses for each ligand[20]. All amino acid ligands form charged interactions between their carboxyl-groups and the N_{δ} group of H464. The diverse side-chain group chemistry of the amino acid ligands is accommodated by a cavernous pocket at the interface between the A- and C-domains.

PDB ID	Isoform	Ligands†	Citation
3SRD	M2	FBP,GOL,K,MG,OXL	Morgan *et al.* (2013)[22]
3U2Z	M2	FBP,DZA	Anastasiou *et al.* (2012)[17]
4JPG	M2	FBP	Guo *et al.* (2013)[56]
4B2D	M2	FBP,SER	Chaneton *et al.* (2012)[19]
4G1N	M2	MG,NZT,OXL	Kung *et al.* (2012)[50]
4RPP	M2	FBP	Wang *et al.* (2015)[57]
4QG8	M2	GOL,K,MG,MLI	Wang *et al.* (2015)[57]
6B6U	M2	B3P,CL,K,MG,OXL,PEG,SCN	Srivastava *et al.* (2017)[58]
4QGC	M2	GOL,K,SO4	Wang *et al.* (2015)[57]
4FXF	M2	FBP,K,MG,OXL	Morgan *et al.* (2013)[22]
4QG9	M2	ACT,MG	Wang *et al.* (2015)[57]
3BJT	M2	MG,OXL	Christofk *et al.* (2008)[23]
1T5A	M2	FBP,GOL,K,MG,OXL,PO4	Dombrauckas *et al.* (2005)[55]
6GG4	M2	CSO,K,PHE,PO4	Yuan *et al.* (2018)[20]
3BJF	M2	FBP,K,MG,OXL	Christofk *et al.* (2008)[23]
4QG6	M2	PRO	Wang *et al.* (2015)[57]

Table 1.1: ***Structures of human PKM2 deposited on the Protein Data Bank.*** †*Ligand acronyms: D-fructose-1,6-bisphosphate (FBP), glycerol (GOL), potassium (K), oxalate (OXL), 1-(2,3-dihydro-1,4-benzodioxin-6-ylsulfonyl)-4-[(4-methoxyphenyl)sulfonyl]piperazine (DZA), L-tartaric acid (TLA), L-serine (SER), N-(4-[4-(pyrazin-2-yl)piperazin-1-yl]carbonylphenyl)quinoline-8-sulfonamide (NZT), magnesium (MG), chloride (CL), malonate (MLI), pyruvate (PYR), di(hydroxyethyl)ether (PEG), L-phenylalanine (PHE), acetate (ACT), sulfate (SO4), phosphate (PO4), S-hydroxycysteine (CSO), unidentified atom (UNX).*

1.2.3 Kinetics of ligand binding and regulation

Allosteric regulation of PKM2 acts over long distances to affect the affinity of the enzyme for its substrate phosphoenolpyruvate. The *apo*-form of the enzyme is reported to have a Michaelis-Menten constant for PEP (K_M^{PEP}) of approximately 2.0 mM[19,55,59]. Activation of PKM2 by FBP is associated with a decreased K_M^{PEP} of 0.2 mM, while the rate of product turnover (k_{cat}) remains unchanged[19,48,55,60]. Some studies, however, have additionally reported an elevated k_{cat} subsequent to FBP binding[20,22,61] and the reason for this discrepancy remains unresolved. The K_M for adenosine diphosphate (ADP; the second substrate) is reported as approximately 0.4 mM and remains unchanged upon allosteric ligand binding[19,55,59].

The reported apparent binding affinity of PKM2 for FBP varies considerably between 210 nM[62] and 7.6 μM[22]. Curiously, isothermal titration calorimetry measurements of FBP binding by Yan *et al.* (2016)[62] found a sub-stoichiometric association of FBP with PKM2, evidenced by an estimated 0.54 binding sites per PKM2 monomer, despite a reported 1.0 stoichiometry found in several FBP-bound PKM2 crystal structures[17,19,23,55]. This result could be explained by the finding that purification of recombinant PKM2 from an *E. coli* expression vector is found to have sub-stoichiometric amounts of co-purified FBP despite extensive dialysis[22,23,63], supporting the notion that FBP binds tightly to PKM2.

Several amino acids including Ser, His, Phe, Ala, Trp, Val and Pro have been shown to regulate the activity of PKM2, though the reported effects of the amino acids on PKM2 enzyme kinetic constants vary. Eigenbrodt *et al.* (1983)[64] first showed that PKM2 was susceptible to pronounced activation by Ser and inhibition by Ala, achieved through a positive or negative effect on the enzyme's affinity for its substrate PEP. More recently, Chaneton *et al.* (2012)[19] reported the binding mode for Ser in the amino acid allosteric pocket with an apparent affinity of 810 μM, acting to decrease the K_M^{PEP} from 1.9 mM to 0.19 mM in the absence and in the presence of saturating amounts of Ser, respectively. Taken together, this led to the prevailing

view that amino acids act as K-type effectors, and that the rate of product turnover does not change upon ligand binding. Nevertheless, a recent contradictory study, reporting four co-crystal structures of PKM2 bound to Ser, Phe, Trp or Ala, purported to show that amino acids have effects on the k_{cat}[20].

In addition to the effects of individual ligands on PKM2 activity *per se*, the multiplicity of distinct allosteric binding pockets on the surface of PKM2 may facilitate concurrent binding of ligands. Therefore, changes to PKM2 enzyme activity are exerted through changes in the concentrations of the metabolite ligands that regulate PKM2. Physical and temporal compartmentalisation of these metabolites may necessitate that PKM2 only responds to specific stimuli, however, it is likely that at any given time several ligands are available to bind to and regulate PKM2.

1.2.4 Concurrent regulation of PKM2 activity by multiple ligands

PKM2 can bind to multiple allosteric ligands *in vitro* that might either reciprocally influence each other's action, or exert independent effects on enzyme activity. Yet, very little is reportedly known about how PKM2 responds to the simultaneous binding of multiple ligands. A single-point mutation at the FBP binding pocket S473Y prevents FBP binding, though this mutant variant can still be activated by Ser[19]. Conversely, H464A abolishes Ser binding but can be activated by FBP[19], suggesting that Ser activation could work independently from FBP activation. Inhibitory amino acids that bind to the same pocket as Ser, however, fail to inhibit enzyme activity in the presence of FBP[18,20]. The apparent dominant effect of FBP activation may suggest that inhibition by amino acids is blocked by FBP binding, through the promotion of an alternate conformation, or that FBP reduces the binding affinities of the amino acids through a cross-talk between the two binding pockets. Similarly, FBP binding can overcome the inhibition of PKM2 by T_3 by sequestering PKM2 away from the monomeric species, to which T_3 binds[24,65].

An alternative explaination for how PKM2 is able to integrate multiple regulatory ligands is, rather than a competitive response, through a synergistic response. This might manifest as two activators (eg. FBP and Ser binding) resulting in supra-activated enzyme activity, or alternatively, an activator and an inhibitor displaying a complex non-competitive kinetic phenotype. In support of the latter, Ashizawa *et al.* (1991)[65] found that FBP addition altered the inhibition of PKM2 activity by Ala, producing only a 30 % inhibition compared to a 97 % inhibition of activity in the absence of FBP. Positive synergism has been experimentally found not to occur between Ser and FBP, and between SAICAR and FBP[19,21]. Interestingly, Zhong *et al.* (2017)[66] recently reported that adenosine monophosphate (AMP) and glucose-6-phosphate (G6P) synergistically activate *M. tuberculosis* pyruvate kinase. The authors found that binding of G6P increased the apparent affinity of the protein for AMP, and conversely, that AMP binding increased the binding affinity of G6P[66]. While the binding of AMP occurs at a pocket equivalent to that of human PKM2 for FBP, G6P binds to a different pocket that is also distinct from the equivalent amino acid interaction site on human PKM2, and it is intriguing to speculate whether a similar cooperative mechanism occurs between FBP and amino acid binding to human PKM2. Moreover, a recent modelling study purportedly found a synergistic effect of concurrent FBP and Ser binding[67]. The authors performed molecular dynamics simulations of monomeric PKM2 and found that the predicted free energy of binding to the substrate PEP changed from -48 kJ/mol (either FBP or Ser bound) to -57 kJ/mol (FBP and Ser). Nevertheless, the free energies reported by Yang *et al.* (2016)[67] translate into a substrate binding affinity of between 10^{-42} M^{-1} and 10^{-35} M^{-1} - a very significant departure from the reported experimental substrate affinity of 1.5 mM. The question of whether binding of allosteric ligands promotes a synergistic effect, therefore, remains unresolved.

The problem of how PKM2 is concurrently regulated is particularly pertinent in an intracellular context, where multiple allosteric ligands co-exist at varying concentrations and bind to PKM2 with a range of affinities. Notably, the intracellular concentrations of the PK cat-

alytic substrate PEP measured in human colorectal carcinoma cells is less than 10 μM and only accumulates to 80 μM upon stable silencing of both PKM1 and PKM2 expression[19]. Low intracellular substrate concentrations suggest that the substrate is limiting for the catalysed conversion into pyruvate. Moreover, the intracellular concentration of FBP has been reported as $\simeq$ 0.5 mM across a panel of cancer cell lines[68], which would represent a 2500-fold excess relative to the reported 200 nM binding affinity to PKM2[62]. Taken together, evidence from the existing literature suggests that concurrent binding of multiple allosteric ligands may be feasible under certain cellular metabolic conditions, and that there is a selective pressure towards a significant fraction of PKM2 remaining constitutively bound to FBP in order to sustain flux through the PK reaction, which is reported to be high relative to other rate-limiting enzymes in glycolysis[69]. Therefore, the interaction between PKM2 and its various allosteric ligands warrants further investigation in order to discern the detailed molecular mechanisms involved in the regulation of this complex glycolytic node, with roles in supporting cancer cell proliferation.

Question 1: To what extent do cellular metabolic conditions exist under which PKM2 is likely to be bound to multiple allosteric effectors? And how are the effects of concurrently bound ligands, with opposing functional signals, integrated by PKM2 to regulate enzyme catalysis?

1.2.5 Regulation of PKM2 structure and oligomerisation

PKM2 molecules form in an equilibrium of monomers and tetramers, with some studies reporting the existence of dimers[62,63,70], in which the tetramer has high enzymatic activity whereas monomers and dimers display basal activity[71]. Many of the PKM2 allosteric ligands have been shown to modulate PKM2 activity by changing the equilibrium between the monomer, dimer and tetramer. The binding of FBP promotes, whereas T_3 prevents, tetrameri-

sation[17,22,62,63,71]. The effect of amino acids on the monomer-dimer-tetramer equilibrium, however, remains controversial. Some reports suggest that Ala promotes the formation of inactive PKM2 dimers[18,70], whereas others have shown that inhibition by Phe, Ala and Trp stabilises an inactive T-state tetramer by forming hydrogen bonds with the carbonyl oxygen atom from R43, and thus repelling the N-terminal helix-loop-helix[20,22].

In addition to ligand-dependent oligomerization, evidence suggests that allosteric ligand binding promotes subtle conformational changes within the tetramer. Crystal structures of *L. mexicana* and *T. cruzi* pyruvate kinase isoforms revealed a rigid-body rocking motion of the A- and C-domains, reorienting them within the tetramer, in the T- to R-state transition[72–74]. Hydrogen-deuterium exchange experiments reported an increase in global flexibility in *E. coli* type 1 pyruvate kinase upon binding of FBP, and proposed that destabilisation of β-strands within the TIM-barrel core resulted from ligand binding[75]. Further attempts to gain insights into the dynamic mechanism of human PKM2 and *M. tuberculosis* pyruvate kinase regulation, found that allosteric activator binding causes R342 on the Aα6-helix to reach across the A-A' interface of the tetramer, forming a stabilising hydrogen bound with the backbone carbonyl group of G271 on the adjacent protomer[22,66].

Ligand-induced conformational changes are likely to be limited to subtle backbone changes and side-chain motions, evidenced by the lack of FBP-driven changes to the radius of gyration of human PKM2 tetramers in small-angle X-ray scattering experiments[62]. A whole-protein mutagenesis study of the human liver isoform of pyruvate kinase (PKL) reported by Tan and Fenton (2017)[76], found that approximately 30 % of residues, when mutated to alanine, perturbed allosteric activation by FBP. The authors concluded that a significant proportion of the protein contributes to allosteric regulation. In the same study, Tan and Fenton (2017)[76] found that perturbing mutations were dispersed across the structure, rather than localised to domain-domain boundaries, suggesting that large scale domain reorientations are not exclusive to the allosteric mechanism of PKL.

The relative contributions of tetramerisation and local structural changes to the mechanism of allosteric activation is disputed. Molecular dynamics simulations of human PKM2 and *L. mexicana* pyruvate kinase have revealed that allosteric activation by D-fructose-1,6-bisphosphate (human) and D-fructose-2,6-bisphosphate (*L. mexicana*) result in the closure of the B-domain over the active site[74,77], consistent with flexible A/ B - inter-domain motions reported in the crystal structure[55,58]. The observation by Gehrig *et al.* (2017)[77] that FBP binding recapitulates some of the experimentally described dynamic changes to the structure, within the monomeric species, argues that the *allosteric content* of the protein is encapsulated within the protomeric unit. Christofk *et al.* (2008)[23] published the first atomic structure of human PKM2 in the *apo*-form, showing that ligand binding is not required for tetramer formation. It is not clear, however, whether apo-tetramer formation is an artefact of the high protein concentration used for crystallisation (80 mg/mL), as polymer self-assembly is inherently a concentration-dependent process. Moreover, PKM2 has been shown to form a mixture of oligomeric states, in solution[22] and in the gas phase[63], though since FBP is known to remain bound to preparations of purified recombinant PKM2, and it is unclear whether the observed tetramers are simply a result of residual FBP.

The apparent dichotomy between protomer-specific conformational and oligomeric effects of allosteric ligand binding on PKM2 activation raises important and fundamental questions about the structural basis of its regulation. In addition to allosteric effectors, covalent post-translational modifications (PTMs) have been suggested to influence the oligomerisation of PKM2 protomers[23,40], although the effects of PTMs on the enzyme mechanism remain elusive. Nevertheless, PKM2 activity in cells is frequently inferred from the oligomeric state that the protein is found to adopt[17,26,40,78–82]. Collectively, current evidence suggests that a link between enzyme activity and the oligomeric state exists and, while not well understood, it may play a role in PKM2 regulation.

Question 2: What are the contributions of ligand-induced conformational and oligomeric changes to PKM2, and what are the relationships between these allosteric effects and PKM2 activity?

Many of the current insights into metabolic dependencies in cancer are derived from analyses of developed tumours and so it is unclear whether observed metabolic changes reflect late adaptive events, or a cumulative agglomeration of metabolic adaptations that have been gradually emerging depending on the needs of cancer cells along the various stages of tumour development[83]. Nevertheless, PKM2, and other allosterically regulated metabolic proteins with roles in disease progression, show that genome-encoded and therefore stably propagated metabolic changes in tumourigenesis are potentiated through protein allostery[83]. Therefore, metabolic flexibility may confer advantages suitable for surviving the varying extracellular conditions cancer cells are subjected to during tumourigenesis, so tools that aid the identification of such mechanisms could hold significant power for discovery.

1.3 Investigating the role of protein dynamics in allosteric regulation

Characterising allosteric mechanisms of enzyme regulation offers several advantages for therapeutic interference with the metabolism of cancer cells. Despite the importance of allostery in regulating numerous aspects of cellular physiology, the investigation of the underlying mechanisms remains a challenge for conventional structural approaches and therefore necessitates multi-disciplinary strategies. A critical obstacle in characterising protein allosteric mechanisms is the inherent flexibility of protein structures, which can result in the emergence of latent allosteric pockets[84,85]. Once an allosteric pocket has been identified, understanding how information is propagated the active site is confounded by the involvement of protein structural motions on a variety of time scales[86]. Furthermore, as exemplified in the case of PKM2, many proteins contain several allosteric pockets that facilitate the simultaneous binding of several ligands. Concurrent binding of multiple allosteric ligands can modulate the functional response of a protein through the action of multiple allosteric pathways[87]. It remains unclear, however, whether such allosteric pathways operate independently, or integrate, either synergistically or antagonistically, to control protein function.

1.3.1 Enzyme regulation can involve structural or vibrational motions

Static model representations of proteins are so prevalent that it is easy to misconstrue these dynamic machines for rigid aperiodic crystals. In addition to the process of protein-folding where a protein reaches its native state through a funnel of energetic intermediates (**Fig. 1.5 A**), folded proteins are soft materials with a complexity, due to the absence of periodicity conditions, forcing the macromolecule to assume a single unique structure. Folded proteins can adopt a variety of different structural and energetic configurations that give rise to an ensemble of isoenergetic ground states, referred to here as conformational *sub-states* (**Fig. 1.5 B**). The amplitudes of the structural motions are determined by the physical and chemical nature of the transient bonded interactions that are broken and made in the transitions between sub-

states, and are characterised by a hierarchy of vibrational and configurational motions (**Fig. 1.5 B**).

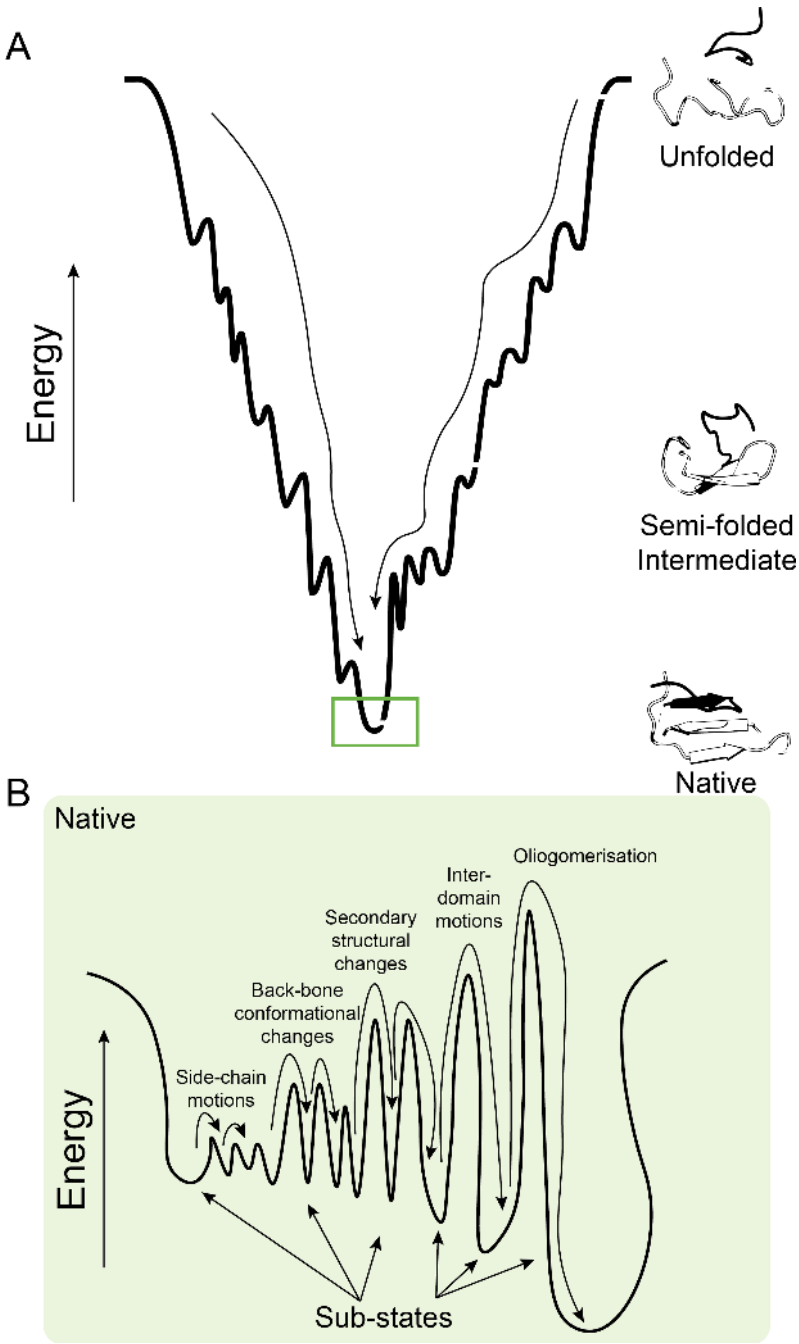

Figure 1.5: ***The energetic landscape of protein folding and native protein dynamics. (A)*** *A cartoon representation of the energetic landscape of protein folding.* ***(B)*** *The native state of a protein consists of an ensemble of conformations, which are termed conformational sub-states. The transitions between conformational sub-states are characterised by a hierarchy of protein dynamics including (i) side-chain motions, (ii) protein backbone conformational changes, (iii) secondary structural changes (iv) inter-domain motions and (v) oligomerisation of multi-meric proteins. Each respective motion necessitates that the protein overcomes energy barriers, the heights of which are determined by the free energy required for a specified motion.*

Initial evidence for conformational sub-states of a protein came in the early 1970s from measurements of the rate of carbon monoxide (CO) re-association to the heme centre of myoglobin (Mb) following photolysis[88–90]. The first crystal structures published in 1960 by Kendrew *et al.* (1960)[91] had shown that Mb does not have a permanent channel through which CO can exit, suggesting that the exit must proceed through a transient channel caused by dynamic fluctuations of the protein. It was therefore hypothesised that the study of ligand escape from its binding to the heme centre as a function of external parameters such as temperature, pressure and solvent viscosity, would yield information on the fluctuations that underpin the protein dynamics of myoglobin[92].

In a seminal paper by Austin *et al.* (1975)[90] the Mb-CO bonded interaction was photolysed with a laser pulse, allowing the CO ligand to migrate within the protein and then finally escape into the bulk solvent. The rate of CO exit (k_{exit}) determined over a range of temperatures is plotted in **Fig. 1.6**, along with typical rate constants of fluctuations of bulk solvent (k_β) and fluctuations of the first hydration shell of a solvated protein (k_α). While the rate of fluctuations in the bulk solvent were found to follow the Arrhenius law, k_α and k_{exit} become non-exponential at temperatures $T < 220K$. This revealed the existence of a glassy state adopted by proteins at very low temperatures, in the absence of protein conformational flexibility. Protein flexibility was found to be restored around 220 K, allowing for an exponential rate of reversible binding of the CO ligand to the heme centre of Mb. The transition of solvated proteins from a glass-like to a soft-matter is thought to be coupled with interactions between the protein and hydration water in the first solvation shell. This explains the observation that the rate of fluctuations of the solvation shell (k_α) undergoes a similar non-exponential to exponential transition around 220 K, as does the rate of CO exit from myoglobin. Subsequent experiments measuring the rate of CO exit under varying solvent compositions, found that the protein must overcome distinct free energy barriers as the CO moves within the protein towards its binding site, providing the first evidence for a hierarchy of protein dynamical motions[88–90]. This theory implied that a molecular system of N components (amino acids)

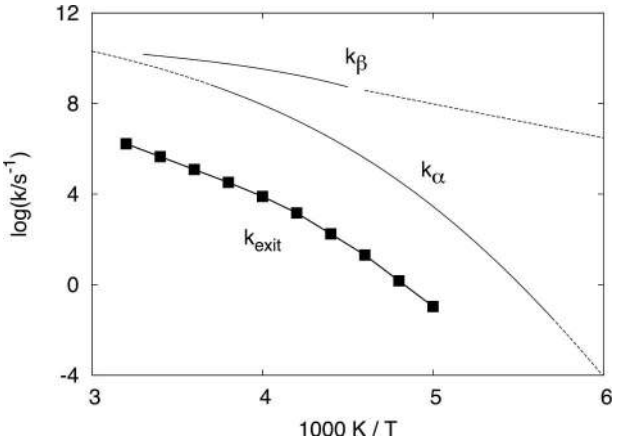

Figure 1.6: ***Typical rate coefficients as a function of temperature.*** *The rate constant k_β rate of fluctuations in the bulk solvent, k_α is the rate constant of fluctuations in the first hydration shell and k_{exit} is the rate constant of CO exit from myoglobin. All rate constants are determined over a range of temperatures. This figure is reproduced from Figure 4 in Frauenfelder et al. (2007)[93].*

has approximately e^N states with energies close to that of the ground state and experiences sizeable fluctuations in its internal energy.

Experimental evidence for the existence of multiple protein sub-states prompted the investigation of the structural-basis of protein dynamics by temperature-dependent X-ray crystallography. Two consecutive papers by Artymiuk *et al.* (1979)[94] and by Frauenfelder *et al.* (1979)[95] provided the first direct experimental evidence for the notion that protein adopt dynamic structures. In the former publication, authors computed the mean-squared displacement of backbone atoms within human lysozyme and found that the residue displacements constitute a characteristic rigid-body motion[94]. Displacements were found to be smaller within the core of the protein and greater on solvent-exposed loops of both lysozyme and myoglobin, due to charge fluctuations in the solvent and the hydration shell[94]. Moreover, Artymiuk *et al.* (1979)[94] found that the catalytic pocket underwent a significant conformational change upon substrate binding, concluding that "...protein mobility may play a significant part in biological activity...".

Allostery had originated from experiments performed by Changeux and co-workers[96–98], and

for decades the two dominant models for allostery were the 'sequential', or KNF (Koshland-Nemethy-Filmer) model[99] and the 'symmetric' or MWC (Monod-Wyman-Changeux) model[98], both of which still contribute to the abstraction of complex allosteric mechanisms[100]. The two essential features of the MWC model were that an equilibrium was proposed to exist between two conformational states, usually known as the R ('relaxed') and T ('tense') states, in the absence of any ligand, and that symmetry is maintained, so that all subunits in a oligomeric protein change between the R and the T states in a concerted manner[101]. Conversely, the sequential model did not require conformational symmetry but instead necessitated a strict application of induced fit, with one conformation existing only when the ligand is not bound[101]. While both the MWC and KNF models agreed on the importance of structural transitions between two pre-defined conformational states whole equilibrium was shifted by the binding of a ligand, both were phenomenological and consequently did not answer the fundamental question of how the binding of a ligand potentiated an allosteric effect at an atomic level of detail[102].

The subsequent accessibility of experimental methods capable of determining protein structure gave rise a stereochemical model of allostery, first proposed by Max Perutz, whereby allostery could be understood in terms of structural changes that could be gleaned through an inspection of high-resolution structures of proteins in the *apo* and *holo* states. In this model, oxygen binding to the T-state triggers the movement of the ferrous iron into the heme plane, realignment of the neighbouring helices and breakage of of inter-subunit salt bridges, thereby shifting the quarternary equilibrium toward the R conformation[103]. The elucidation of the initial structures of haemoglobin in two forms (the T- and R-forms) supported the notion that allostery as a general phenomenon might be understood, perhaps in all systems, if only the structures of the allosteric states and the transitions between those states could be determined. Nevertheless, crystal structures are limited in that they provide only a limited snapshot of a single *consensus* conformational state.

Moreover, despite the fact that plots of estimated mean-square fluctuations versus residue

number remain a standard part of the analysis of high-resolution structures, the contribution of overall translation and rotation as well as crystal disorder to B-factors, continues to be a source for concern in their interpretation[104]. It was understood that the best means to study the structural basis of cooperativity displayed by haemoglobin, and other allosteric proteins, was to have available a way of calculating the energy of the protein as a function of the atomic positions. To this end, Gelin and Karplus developed a method for calculating the side-chain potentials of bovine peptide tripsin inhibitor (BPTI) using an empirical energy function[105]. The same approach was applied towards calculating the forces on the atoms of haemogloblin for the minimisation of its energy in response to ligand binding on the heme group[106], providing the first comprehensive description of distal side-chain motions involved in the transition.

Next, Karplus and co-workers used the forces computed from an empirical potential energy function in Newton's equation to simulate protein dynamics, which came to be known as molecular dynamics (MD) simulations. MD simulation methods had already been developed, prior to the 1970s, for simpler systems such as liquid argon with a soft-sphere (Lennard-Jones) potential[107], followed by simulations of complex fluids[108]. The first MD simulation of bovine peptide tripsin inhibitor (BPTI) *in vacuo* performed by McCammon *et al.* (1977)[109], made clear that proteins are relatively soft polymers and, consequently, have significant structural fluctuations at room temperature; i.e., the static view of the structure of biomolecules had to be replaced by a dynamic picture.

While allosteric transitions are often accompanied by structural changes, long-range communication between sites can also be mediated by changes in dynamic fluctuations about the mean conformational state (ie. in the absence of structural changes), which are driven by entropic changes in the protein. Dryden and Cooper (1984)[110] first proposed that an allosteric effector may simply change the broadness of the free energy basin of the protein conformational state, rather than shifting the basin to a distinctly different region. This provided a plausible model for an entropy-driven allosteric mechanism, in the absence of stereochemical

conformational changes. Entropy-driven allosteric mechanisms have since been experimentally described. Popovych *et al.* (2006)[111] characterised the negatively cooperative binding of cAMP to dimeric catabolite activator protein (CAP). Two cAMP molecules bind to dimeric CAP with negative cooperativity and function as allosteric effectors by increasing the protein's affinity for DNA. The authors showed that the binding of the first cAMP molecule had minimal effects on the conformation of the second cAMP binding site, but rather that distinct change in protein motions between the two sequential cAMP binding steps results in a large difference in conformational entropy[111]. This showed that allostery can be mediated exclusively by vibrational changes, rather than large-scale conformational changes.

1.3.2 Allostery is an ensemble phenomenon

Just as protein folding can be viewed as a landscape of possible conformational states surrounding the *native* (free energy minimum) state, a folded protein has an ensemble of conformational sub-states. According to this statistical framework, all possible conformations of a protein are populated within the ensemble according to their free energies. The free energy landscape of a folded protein can be smooth with many accessible states, rough with only a few states, or somewhere in between[112]. Allosteric ligands can effectively remodel the energy landscape of proteins, manifesting as enthalpic and/or entropic redistributions within the available configurational space. Therefore, protein function is not determined purely by its static structure but rather through a redistribution of already existing populations in response to ligand binding.

The energy landscape of a folded protein contains a plurality of functional mechanisms and, as such, sub-states within the ensemble can be partitioned into configurations that account for the active and inactive states of the enzyme. Nevertheless, investigating the relationships between protein dynamics, activity and structure is hampered by the difficulty in observing all three characteristics simultaneously. A landmark study by Volkman *et al.* (2001)[113], used nuclear magnetic resonance (NMR) spectroscopy to correlate the structural states of

the nitrogen regulatory protein C (NtrC) and its interconversion dynamics directly with its biochemical activity. Phosphorylation of NtrC causes the displacement of several structural elements resulting in the exposure of a hydrophobic patch on one side of the molecule, enabling NtrC dimers to self-associate to form oligomers that regulate the transcriptional activity of a number of genes involved in nitrogen metabolism in enteric bacteria. In order to probe μs - ms time-scale dynamics, Volkman *et al.* (2001)[113] measured the chemical shifts of protein backbone atoms and found that the dynamics of various NtrC mutants correlated with their activity, supporting the idea of a population shift between two pre-existing conformations.

The propensity of an allosteric transition to stabilise a particular sub-state within the conformational landscape, and hence the probability that a protein will exert a given function, is determined by the free energy barrier between the different sub-states. Indeed, within the boundaries of the protein's primary sequence, the free energy between all of the accessible sub-states constrains the landscape of a protein and encodes the structure and function. Yao *et al.* (2008)[114] used solution NMR measurements to quantify the free energy of the light-sensitive LOV2 (light, oxygen and voltage) domain-Jα-helix binding equilibrium following light stimulation. Blue light absorption of the LOV domain induces the formation of a covalent adduct between a conserved cysteine residue in the LOV domain and the C4a carbon of the isoalloxazine ring of FMN, leading to a light-dependent enhancement of phototropin kinase activity. Yao *et al.* (2008)[114] found that this light-sensitive switch is mediated by an equilibrium between two conformations, determined by the positioning of the Jα helix between a bound conformation (dark; inactive) and a dissociated state (lit; active). NMR relaxation experiments found that photo-excitation resulted in a redistribution between the bound-dissociated Jα equilibrium from 98:2 in the dark state to 9:91 in the light state, with an associated free energy difference of 3.8 kcal mol^{-1}. The authors noted that the free energy barrier of the allosteric transition was small relative to the energy of the blue light photons being absorbed by the flavin chromophore ($\simeq$ 64 kcal mol^{-1}). Thus, the bonded and non-bonded interactions that make up the protein core fold are generally modest, suggesting that no single state will

necessarily dominate the ensemble. As such allosteric mechanisms are likely to be controlled by statistical equilibria and less deterministic than was suggested by classical discrete-state models.

As exemplified by NtrC and LOV2, allosteric regulation of proteins can involve transitions between sub-states with unique physical and biochemical properties. Therefore, the role that structural dynamics of a protein play in biological processes can only be understood by characterizing all thermally accessible protein conformations and their populations. Since the free energy landscape of a folded protein contains many conformational sub-states, any macromolecular observable A can be described as an ensemble average over microscopic sub-states given by:

$$A = \langle a \rangle_{ensemble} = \frac{1}{N} \sum_{\lambda=1}^{N} a(x_\lambda) \tag{1.2}$$

where x_λ is the sub-state of the λth member of the ensemble. In reality, experimental measurements are made only on a single system and all the microscopic detailed motion is present. As such, one observes an averaged observable over time of the detailed motion, which suppresses the microscopic details. Thus, the time average and the ensemble average should be equivalent:

$$A = \langle a \rangle_{ensemble} = \lim_{t \to \infty} \frac{1}{t} \int_0^t dt\, a(x(t)) \tag{1.3}$$

It is often straightforward to determine $\langle a \rangle$ from an experiment, while the underlying distribution of $a(x)$ is experimentally inaccessible since the experiment is a time- and ensemble-average over molecular conformations[115]. A common approach to dynamic interpretation of NMR relaxation experimental data is to use NOE data and s^2 order-parameters as constraints for molecular dynamics simulations to simulate a conformational ensemble in agreement with experimental data[116]. Such approaches are limited in their application due to complications in

performing measurements in a sufficiently diverse combination of alignment media[117]. Nevertheless, molecular dynamics simulations can provide a direct sampling of microscopic sub-states $a(x)$ and, as such, present a powerful tool for investigating allosteric transitions of proteins.

1.3.3 Identification of allosteric pathways from molecular dynamics simulations

Empirical evidence suggests that subsequent to ligand binding to an allosteric site, a network of residues mediates the communication between the allosteric site and the functional centre of the protein (catalytic pocket in the case of enzymes). Communication across the protein, through so-called allosteric pathways[87], is mediated by networks of residues that exhibit spatial and/or temporal correlations. It was therefore hypothesised that measuring correlated motions within a protein structure during allosteric transitions, observed from simulated trajectories of the protein conformational ensemble, would lead to the identification of the residues involved in propagating information along allosteric pathways.

Covariance measurements of residue correlation

Long-range positional correlations were first observed in molecular dynamics simulations of bovine trypsin inhibitor (BTPI) and hen white lysozyme (HEWL) by Huenenberger *et al.* (1995)[118]. The authors measured correlated motions by computing the normalised covariance matrix of atomic fluctuations, after a superposition of the trajectory to the first frame:

$$C_{ij} = \sqrt{\frac{\langle x_i \cdot x_j \rangle}{\langle x_i^2 \rangle \langle x_j^2 \rangle}} \tag{1.4}$$

where x_i and x_j are the vectorial positional fluctuations of atoms i and j, respectively. As noted by Ichiye and Karplus (1991)[119] correlations measured of the form in Equ. 1.4 assumes that x_i and x_j are co-linear vectors, since C_{ij} depends on the angle between both vectors. Moreover, the use of the covariance matrix implies a Gaussian approximation of the underlying configurational space density, and therefore this approach treats correlations in a quasi-harmonic

approximation[120]. The quasi-harmonic approximation of the Hamiltonian implies that the potential energy surface of a protein is parabolic, which is irreconcilable with empirical findings showing that proteins have an ensemble of conformational sub-states (Section 1.3.2).

Mutual information

Rather than measuring co-linear motions, mutual information presents a more physically faithful method for computing correlated motions from protein dynamics. The fluctuations of each atom are considered to have a given distribution, and the correlations between distributions are calculated by the joint probability distribution. The fluctuations of two atoms are considered to be completely independent if their joint probability distribution is equal to their marginal probabilities:

$$p(\mathbf{x}) = \prod_{i=1}^{N} p_i(x_i) \tag{1.5}$$

where $p(x)$ is the canonical ensemble density $p(x) = Z^{-1}exp\left(-\frac{1}{k_B T}V(x + \langle r \rangle)\right)$, with partition function Z, temperature T, Boltzmann constant k_B, potential energy V and marginal probability density $p_i(x_i)$. The idea behind this analysis to calculate those correlations that violate Equ. 1.5 using the Shannon mutual information:

$$I[x_1, x_2, ..., x_{3N}] = \sum_{i=1}^{3N} H[x_i] - H[\mathbf{x}] \tag{1.6}$$

where H is the entropy of the random variables defined as: $H[\mathbf{x}] = -\int p(\mathbf{x})\ log\ p(\mathbf{x})\ d\mathbf{x}$. Focusing on correlations between pairs of atoms:

$$I[\mathbf{x_i}, \mathbf{x_j}] = H[\mathbf{x_i}] + H[\mathbf{x_j}] - H[\mathbf{x_i}, \mathbf{x_j}] \tag{1.7}$$

The mutual information between pairs of fluctuating atoms was first used by Lange and Grubmueller (2006)[120] to measure correlated motions from molecular dynamics simulations of the

B1 domain of G protein and T4 lysozyme. The authors noted that a linear covariant qualification of the covariance matrix missed more than 50 % of the correlations, attributed to a dependence on mutual orientations of the atomic fluctuations and non-linear correlations that emerged in the dynamics[120]. As such, linear mutual information between distal protein motions has subsequently emerged as a commonly used method for computed correlated motions[121,122], and is not susceptible to the problematic assumptions of covariance measures previously highlighted.

Discrete-state mutual information

More recently, the use of mutual information to measure correlated motions that underpin allosteric communication was further developed by Kleinjung, Fraternali, and co-workers[123,124] using a discrete-state formalism for the mutual information calculation. The authors defined a collection of 24 four-residue fragments, which comprised the so-called M32K25 *structural alphabet*[125]. Each fragment provides a simple and explicit description of four successive Cα atoms along the backbone of a protein and defined by three internal angles. While limited to describing positions of Cα atoms, and not considering other fine-grained features of protein structure, the M32K25 structural alphabet (SA) was effective in encoding the structure of all experimental protein structures deposited in the Protein Data Bank, and could therefore be used as a coarse graining method to define the local orientation of a protein backbone in a simulated ensemble (**Fig. 1.7 A**). In this context, the *local orientation* of a protein backbone is defined as the vectorial position of successive Cα atoms within a protein structure, which is discretised into overlapping four-residue long segments. Insofar as the M32K25 SA maps the vectorial position of successive Cα atoms, and thereby changes to the positions of these segments that result from simulated or experimentally obtained protein dynamics, the detected changes are localised to the affected regions of the protein backbone. This makes the analysis of a molecular dynamics trajectory with the M32K25 SA particularly sensitive to backbone conformational changes (e.g. secondary structure transitions) and rather insensitive to side-chain fluctuations. Rather than a vectorial definition of atomic fluctuation, the discrete-state

mutual information between locally-encoded fragments was computed in order to determine distal correlations across the structure:

$$I^n(C_i;C_j) = \frac{I(C_i;C_j) - \epsilon(C_i;C_j)}{H(C_i,C_j)} \tag{1.8}$$

where the columns of the structural fragment alignment are given by C_i and C_j, $I(C_i;C_j)$ is the mutual information, $\epsilon(C_i;C_j)$ is the expected finite size error and $H(C_i,C_j)$ is the joint entropy. The mutual information is given by

$$I(C_i;C_j) = \sum\sum p(c_i,c_j)\, log_2\, \frac{p(c_i,c_j)}{p(c_i)p(c_j)} \tag{1.9}$$

where the two columns in the structural alphabet alignment C_i and C_j are random variables with a joint probability mass function $p(c_i,c_j)$, and marginal probability mass functions $p(c_i)$ and $p(c_j)$. The joint entropy $H(C_i,C_j)$ is defined as

$$H(C_i;C_j) = -\sum\sum p(c_i,c_j)\, log_2\, p(c_i,c_j) \tag{1.10}$$

The discrete mutual information calculated for finite state probabilities can be significantly affected by random and systematic errors. In order to account for this, an error term $\epsilon(C_i;C_j)$ was subtracted from the mutual information $I(C_i;C_j)$ in Equ. 1.8 given by

$$\epsilon(C_i;C_j) = \frac{B^*_{C_iC_j} - B^*_{C_i} - B^*_{C_j} + 1}{2N} \tag{1.11}$$

where N is the sample size and $B^*_{C_iC_j}$, $B^*_{C_i}$ and $B^*_{C_j}$ are the number of states with non-zero probabilities for C_iC_j, C_i and C_j, respectively.

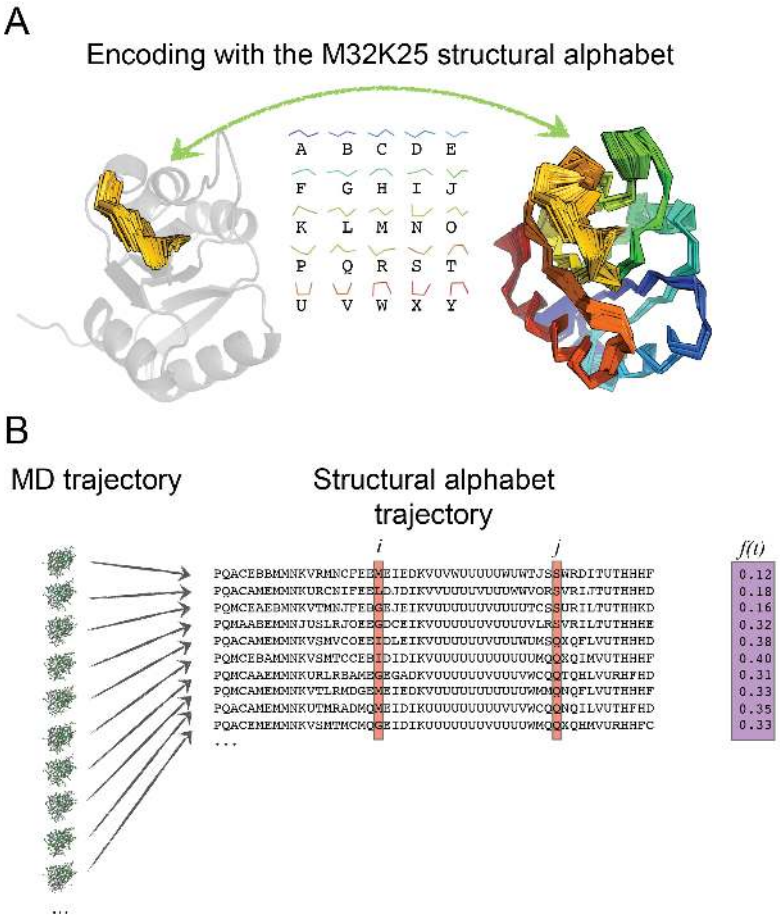

Figure 1.7: ***The M32K25 structural alphabet as a fragment-based representation of protein structure and dynamics.*** ***(A)*** *An illustration of the fragment-encoding procedure, whereby a protein is encoded with the M32K25 structural alphabet. Each sequential four-residue stretch is encoded with one of 25 fragments, previously defined by Pandini et al. (2010)*[125]*.* ***(B)*** *Multiple conformational snapshots, derived from a molecular dynamics trajectory, are encoded with a string of fragments giving a multiple sequence alignment of the conformational state of each snapshot in the trajectory. The mutual information between each fragment in the protein is then determined by calculating the marginal probabilities and the paired probabilities for each combination of columns in the alignment. This figure is partially reproduced with permission of Dr. Alessandro Pandini.*

Pandini *et al.* (2012)[123] used the M32K25 structural alphabet to encode the ensemble of structures calculated from molecular dynamics simulations of the bacterial two-component system NtrC, as a basis for computing the mutual information between distal fragments in the protein. The authors found that the transmission from the phosphorylation site to the signalling surface of the receiver domain NtrC was mediated by a layer of hub residues[123]. Moreover, the location of the hubs preferentially connected to the allosteric site was found to be in close agreement with key residues experimentally identified as involved in the signal transmission.

A recent investigation of the allosteric mechanism of a myosin small-molecule activator omecamtiv mecarbil (OM) binding to cardiac myosin, used the discrete-state mutual information approach developed by Pandini *et al.* (2012)[123] to investigate the mechanism of action of myosin activation. Hashem *et al.* (2017)[126] found that OM binding to an allosteric pocket both increased coupling of the motions of the converter and lever arm subdomains to the rest of the protein, and established allosteric communication pathways between the OM binding site and the functional regions in the U50K subdomain. This case-study, and others[123,124,127–130], demonstrates the utility of discrete-state mutual information as a predictor of allosteric pathways.

An important distinction between the method developed by Pandini *et al.* (2012)[123] and the linear mutual information method developed by Lange and Grubmueller (2006)[120], is that the Pandini *et al.* method normalises the mutual information between fragments by the combined entropies. Determining the positional entropies of fragment-encoded structures is trivial with a discrete-state model (see Equ. 1.10), and can therefore easily be computed over each fragment in the structural alignment. Conversely, the method by Lange and Grubmueller calculates the mutual information in Cartesian space. Therefore, it would be necessary to assume the quasi-harmonic approximation in order to compute the positional entropy, offsetting the advantages of mutual information as a statistical framework. Normalising for the combined

entropy of each combination of fragments has the presumed effect of reducing false-positives by normalising for random entropic motions, though a robust comparison between the two methods is lacking. Nevertheless, due to the ease of normalising the correlations for random thermal fluctuations within a discrete-state model, the method developed by Pandini *et al.* (2012)[123] holds particular promise for revealing allosteric communication pathways.

Computer simulations of protein dynamics, including molecular dynamics (MD) stochastic dynamics (SD) and Monte Carlo (MC) simulations, continue to provide a powerful toolbox for the investigation of allosteric transitions. An ensemble view of protein allostery[86] predicates that proteins sample an ensemble of conformations and that the configurational landscape is modified by allosteric ligand binding. This landscape reconfiguration is achieved by pathways of correlated residues, which can be accurately determined by computing the mutual information between distal residues. Nevertheless, each minima in the configurational landscape accessible to a protein gives rise to unique structural and dynamic properties, and therefore measurements of correlated motions should account for this sampling of phase space, though no methods to this end exist. Given the power of trajectory methods for elucidating allosteric transitions, the present work makes use of molecular dynamics simulations to probe the conformational dynamics of PKM2.

Question 3: To what extent can networks of correlated motions, extracted from defined conformational sub-states, be used as a means of predicting protein residues that propagate the energetic effects of allosteric ligands?

1.4 Summary of research questions and the contribution of this thesis

Understanding the molecular basis for how PKM2 catalytic activity is regulated is critical towards investigating its role as a disease target. This thesis follows on from work by Christofk *et al.* (2008)[26], Anastasiou *et al.* (2011)[40] and Anastasiou *et al.* (2012)[17] detailing how reduced PKM2 catalytic activity favours pro-tumorigenic cell growth. The prevailing view that allosteric regulation of PKM2 promotes its disease-associated functions was subsequently supported by the a number of compound screens that led to the identification of small molecule activators of PKM2, which were found to alleviate tumour progression by stabilising the highly active tetrameric form of PKM2[17]. Additional work by Chaneton *et al.* (2012)[19], and others, established a cross-talk between glycolysis and *de novo* serine biosynthesis, orchestrated by allosteric regulation of PKM2 by serine and SAICAR. Moreover, work by Morgan *et al.* (2013)[22] and more recently by Yuan *et al.* (2018)[20] found that PKM2 acts as a nutrient sensor for a number of amino acids, which can activate or inhibit enzyme activity by binding to a common allosteric pocket. Taken together, these studies led us to hypothesise that the cellular activity of PKM2 is regulated by numerous allosteric ligands, and that this may be relevant for metabolic phenotypes associated with cancer cell growth. Allosteric regulation of PKM2 activity by numerous endogenous ligands may occur either in isolation; or concurrently, resulting from the multiplicity of distinct allosteric pockets on the surface of PKM2. While the mechanisms of how PKM2 responds to ligands *per se* has been carefully investigated, very little is understood about how PKM2 responds to simultaneous binding of multiple allosteric effectors with opposing functional signals. This led to the first research question: *Do cellular metabolic conditions exist under which PKM2 is likely to be bound to multiple allosteric effectors? And if so, how are the effects of concurrently bound ligands, with opposing functional signals, integrated by PKM2 to regulate enzyme catalysis?* To this end, Chapter 3 investigates possible functional mechanisms of allosteric regulation by a number of ligands, both alone and in combination, on PKM2 enzyme activity.

Regulation of PKM2 enzyme activity has been suggested to involve oligomeric changes and possible subtle conformational rearrangements. Early investigations of PKM2 regulation by Hofmann *et al.* (1975)[70] suggested that the oligomeric structure of the protein can alternate between lower-order oligomers and tetramers. Subsequent work by Kato *et al.* (1989)[24], Anastasiou *et al.* (2012)[17], Morgan *et al.* (2013)[22] and Yan *et al.* (2016)[62] found that stabilisation of PKM2 tetramers resulted in high enzyme activity. Nevertheless, the monomer-dimer-tetramer distribution of PKM2 in the absence of any allosteric ligands, and the subsequent effects of various ligands on this distribution, is disputed. Moreover, if multiple ligands concurrently regulate PKM2, under cellular metabolic conditions, the necessary relationship between enzyme activity and oligomerisation remains to be determined. It may, for example, be the case that a molecule of PKM2 is bound to both FBP (an activator; reportedly promotes tetramerisation) and alanine (an inhibitor; reportedly destabilises tetramers). Given the current available knowledge in the literature, it is unclear how ligands alone, and in combination, determine the oligomeric state of PKM2, and whether this is linked to the prevailing level of enzyme activity. This led to the second research question, which will be addressed herein: *What are the contributions of ligand-induced conformational and oligomeric changes to PKM2, and what are the relationships between these allosteric effects and PKM2 activity?* The relationship between PKM2 activity and its oligomeric state is of greater biological interest given that PKM2 activity in cells is often inferred from the its oligomeric state[17,26,40,78,79,81,82]. To this end, Chapter 4 presents an investigation of the oligomeric and structural conformation of PKM2, and how this changes in response to concurrent allosteric ligand binding, using native mass spectrometry and ion mobility. Chapter 5 builds on the experimental findings of the previous chapter, and presents details about the conformational dynamics of PKM2 in response to allosteric regulation, gleaned from molecular dynamics simulations.

The previous finding by Anastasiou *et al.* (2012)[17] that allosteric activation of PKM2 alleviates some of its pro-tumourigenic functions, suggests that genome-encoded, and therefore

stably propagated, allosteric effects can contribute towards disease progression. Allosteric pockets are subject to less evolutionary pressure and hence are more sequence variable than orthosteric sites[131], affording exogenous allosteric ligands greater target specificity. Therefore, methods which [i] identify putative allosteric pockets and [ii] elucidate networks of protein residues involved in the propagation of allosteric signals, would hold great power for drug discovery. Particularly successful has been the use of analysis methods such as those developed by Pandini *et al.* (2012)[123] and Lange and Grubmueller (2006)[120] used to identify distally correlated motions in the backbone of proteins in order to elucidate allosteric pathways. Nevertheless, work by Volkman *et al.* (2001)[113], Lindorff-Larsen *et al.* (2005)[116], Yao *et al.* (2008)[114], Guerry *et al.* (2013)[117] and many others, has shown that protein dynamics results in an ensemble of pre-existing conformations that determine the physico-chemical and functional properties of enzymes. In this respect, ensemble-averaging of any dynamic time-resolved observable is required. Additional methods are required to predict the network of residues involved in protein allostery, with a consideration for the ensemble nature of protein dynamics. This led to the third and final research question: *To what extent can networks of correlated motions, extracted from defined conformational sub-states, be used as a means of predicting protein residues that propagate the energetic effects of allosteric ligands?* To address this question, Chapter 5 presents the development of a *AlloHubMat*, a novel method and stand-alone software package used to predict allosteric hub residues from molecular dynamics simulations using tools from information theory. AlloHubMat is applied towards the analysis of molecular dynamics simulations of PKM2 and predicts a network of residues involved in propagating the effects of FBP-induced activation. These *in silico* predictions are tested in Chapter 6 using single-point mutant variants of PKM2, designed to abrogate the allosteric pathways induced by FBP activation.

In addition to various biophysical and analytical methods, molecular dynamics simulations are used as a detailed statistical-mechanical technique for investigating the conformational dynamics of PKM2 in response to allosteric regulation. Moreover, the simulated trajectories

generated from MD calculations are used as an input for AlloHubMat, to predict allosteric pathways. Given the importance of molecular dynamics as an *in silico* tool for this work, and the immense power it presents for the study of protein dynamics, the mathematical and physical principles of MD simulations are detailed in the following Chapter.

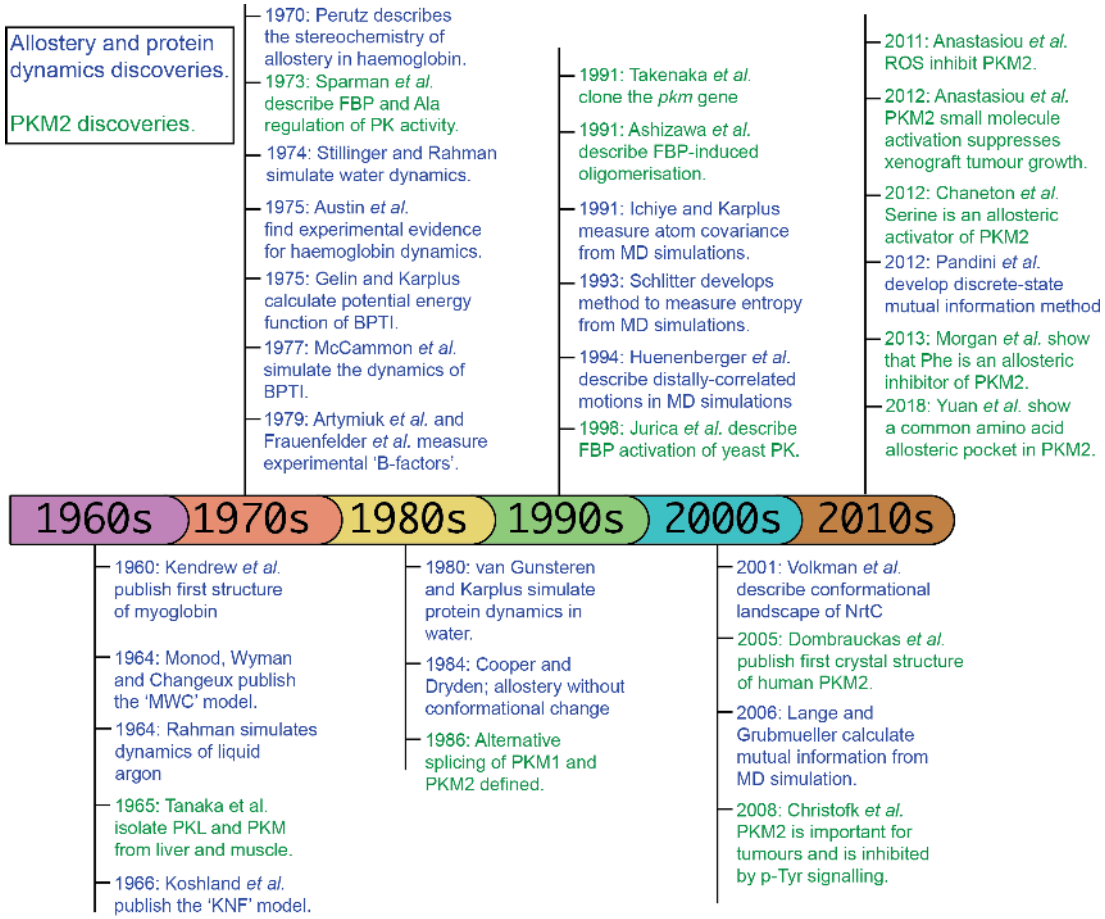

Figure 1.8: ***Time line of historical discoveries in the fields of protein dynamics and allostery, and PKM2 biology and biophysics.*** *Historical discoveries relating to the fundamental concepts of protein dynamics and allostery, with a focus on molecular dynamics simulation methods, are shown in dark blue. Discoveries relating to PKM2 biology, structure and regulation are shown in dark green.*

Chapter 2

Theoretical, computational and experimental methods

2.1 Theoretical principles of molecular dynamics simulations

> *[Anfinsen] showed a film of the folding of a protein with "flickering helices forming and dissolving and coming together to form stable substructures". Of course, the film was purely imaginary, but it led to my asking him whether he had thought of taking the ideas in the film and translating them into a quantitative model. He said that he did not really know how he would do this, but to me it seemed clear that such a model could be based on straightforward physical concepts.*[132]
>
> *Martin Karplus*

The aim of a molecular dynamics (MD) simulation is to compute the equilibrium and transport properties of a classical many-body system. The term classical means that the interaction between particles and their movement is studied over a given time t through a classical statistical mechanical treatment of the constituent particles. The real power of MD simulations in studying protein dynamics is in the ability to complement experimental studies. MD simulations can act as a figurative *microscope* affording the investigator high spacial and temporal resolution

in the study of macromolecular dynamics. Given that most current applications of molecular dynamics simulations of large proteins use a classical statistical mechanical framework due to the enormous computational cost associated with solving the time-dependent Schroedinger equation for biologically interesting molecules, the Hamiltonian of a classical system will be introduced in the following section. This will be followed by a discussion of how molecular forces are computed in an MD simulation, a consideration for how dynamics is calculated in difference physical ensembles and finally, a brief description of the potential energy function used to represent the atomic interaction energies in an MD simulation.

2.1.1 The Hamiltonian of a classical mechanical system

The Hamiltonian (H) can be used to describe a classical system of particles i with coordinates $\mathbf{r}$ and momenta $\mathbf{p}$. The Hamiltonian is equal to the total energy if the potential energy function is independent from time and velocity:

$$H(\mathbf{r},\mathbf{p}) \equiv K(\mathbf{p}) + U(\mathbf{r}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}) \tag{2.1}$$

where K is the kinetic energy, U is the potential energy and m is the mass. A classical system can therefore be defined by the set of values $\mathbf{r},\mathbf{p}$, which corresponds to a point in phase space defined by the internal coordinates of the system and their momenta.

In order to measure thermodynamic averages over a micro-canonical ensemble of states characterised by the macroscopic variables (N,V,E), it is necessary to determine the probability distribution of finding the system at each point in phase space ($\rho(\mathbf{r},\mathbf{p})$). With a knowledge of this distribution of states, one can compute the phase space average of any dynamic variable ($A(\mathbf{r},\mathbf{p})$) of interest:

$$\langle A \rangle = \frac{tr\left(e^{-\beta H(\mathbf{r},\mathbf{p})}\right) A}{tr\left(e^{-\beta H(\mathbf{r},\mathbf{p})}\right)} \tag{2.2}$$

where H is the Hamiltonian of the system, $\beta = \frac{1}{k_B T}$ and tr is the trace of the operator. With this it is assumed that each quantum state of a many-body system with energy E is equally likely to be occupied. The same thermodynamic observable can be calculated for all quantum states of a system i:

$$\langle A \rangle = \frac{\sum_i e^{-\beta E_i} < i|A|i >}{\sum_j e^{-\beta E_j}} \tag{2.3}$$

Unfortunately computing thermal averages for a many-body system by solving the Schroedinger equation and then computing the expectation value of the operator A for all quantum states that have a non-zero statistical weight is not possible for macromolecular systems. The number of quantum states contributing to the operator A in Equ. 2.3 would be so large ($10^{10^{23}}$) that a numerical evaluation of the system would be impractical. For this reason the basic assumption of statistical mechanics is made, whereby it is assumed that a system with fixed number of particles (N), volume (V) and energy (E) is equally likely to be found in any of its quantum states. Such an average over all quantum states of an N-body system in Equ. 2.2 is called an *ensemble average*. Nevertheless, in order to compare an observable computed from a simulation with that derived from an experiment, it is desirable to study the time-evolution of that system. Thermal averages can be computed by following the motion of the system through the phase space as a function of time (t) by integrating the system's equations of motion, taking the averages only over those points that are visited along the trajectory. Provided that the initial coordinates and momenta of all atoms are defined ($\mathbf{r}^N, \mathbf{p}^N$), the time-averaged density ($\overline{\rho_i(r)}$) can be measured in a molecular dynamics simulation in a volume V, at a constant total energy E:

$$\overline{\rho_i(r)} = \lim_{t \to \infty} \frac{1}{t} \int_0^t dt' \, \rho_i(r; t') \tag{2.4}$$

The ergodic hypothesis states that when t tends towards infinity and that the trajectory samples the entire phase space, the time average does not depend on the initial conditions.

Therefore, averaging over all initial phase space coordinates is equivalent to averaging over the time-evolved space coordinates:

$$\overline{\rho_i(r)} = \langle \rho_i(r) \rangle_{NVE} \tag{2.5}$$

Generally the above is not true due to the finite length of molecular dynamics simulations, which is constrained by computer processing speed. As such, ergodicity is unattainable for simulations of macromolecules in the vast majority of cases. The ideal simulation length depends primarily on the time-scale of the phenomena of interest. Taking the intrinsic dynamic motion of protein atoms as an example, different atomic and molecular motions are dispersed across a wide range of time scales. The positions of amino acid side chains may fluctuate with a relatively with relatively high frequency (on the ps time scale), whereas large inter-domain tertiary structural rearrangements of the entire protein may occur on much longer time scales (approaching the μs to ms time scales). In reality, the time scales of these motions largely varies from protein to protein and may depend on the physical and chemical properties of the polypeptide chain. Nevertheless, slow structural dynamics and hence long time scales, still represent a significant computational cost. Apart from using a 'brute-force' approach, this can be overcome by decreasing the number of degrees of freedom of the system, or by biasing the energetic landscape such that the protein is able to traverse free energy barriers and explore conformational space otherwise inaccessible. Neither of these approaches are the subject of this thesis and, as such, will not be reviewed here.

2.1.2 The force calculation and integrating the equations of motion

To start a molecular dynamics simulation, it is first necessary to assign the initial positions and velocities to all particles of the system. Following this initialisation step, the forces on each particle within the system can be calculated and Newton's equations of motion can be solved. The positions of the atoms are adjusted such that steric clashes between the composite particles are minimised. Additionally, prior to the start of the simulation, velocities

are assigned to each particle such that the total momentum is zero and that the instantaneous temperature T matches a desired value:

$$k_B T(t) \equiv \sum_{i=1}^{N} \frac{m v_i^2(t)}{N_f} \tag{2.6}$$

Next, the forces acting on each of the particles in the system are computed. The forces are subsequently used, along with the positions of the particles at the current and at the previous step, in order to predict the positions at the next step. If a given pair of particles are close enough to interact, the forces along the x-direction between these particles are derived from the potential energy ($U(\mathbf{r}^N)$):

$$f_x(\mathbf{r}) = -\frac{\partial U(\mathbf{r})}{\partial x} \tag{2.7}$$

The potential energy between each interacting particle in the system is defined by empirically-defined force field potentials. The functional form of these potentials will be described later. It is worth mentioning here that the force evaluation is the most time-consuming step of a molecular dynamics simulation, as it is necessary to consider the contribution to the force on particle i due to all its neighbours. Considering only the interaction between a particle and the nearest image of another particle, it is necessary to evaluate $N \cdot \frac{N-1}{2}$ pair distances, which scales as N^2.

Having computed all the forces for each of the constitutive particles of the system, Newton's equations of motion can be integrated. The simplest way to construct an integrator is through a Taylor expansion of the positions and velocities:

$$r(t + \Delta t) = r(t) + v(t)\Delta T + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta^3}{3!}\dddot{r} + O(\Delta t^4) \tag{2.8}$$

$$r(t - \Delta t) = r(t) - v(t)\Delta T + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta^3}{3!}\dddot{r} + O(\Delta t^4) \tag{2.9}$$

$$r(t+\Delta t) \simeq 2r(t) - r(t-\Delta t) + \frac{f(t)}{m}\Delta t^2 \tag{2.10}$$

These equations have been modified to make them time-reversible[133], thus increasing the accuracy and robustness of the integration.

2.1.3 Molecular dynamics in various ensembles

Until this point, concepts for simulating a system in the microcanonical (NVE) ensemble have been introduced. Nevertheless, in order to replicate the conditions of a particular experimental set-up, it is often desirable to perform molecular dynamics calculations in different ensembles (ie. NVT or NPT). Physical and mathematical considerations for simulating in these ensembles will be discussed briefly.

Constant temperature (NVT)

A constant temperature can be imposed by bringing the system in thermal contact with a large heat bath. This allows for the study of different molecular systems at different temperatures and sampling of the canonical statistical ensemble. Under these conditions, the probability of the system populating a given energy state is given by the Maxwell-Boltzmann velocity distribution:

$$P(p) = \left(\frac{\beta}{2\pi m}\right)^{3/2} \cdot exp\left[\frac{-\beta p^2}{2m}\right] \tag{2.11}$$

We then obtain a simplified relation between the imposed temperature T and the kinetic energy for each particle in the system

$$k_B T = m\langle v_\alpha^2 \rangle \tag{2.12}$$

where k_B is the Boltzmann constant, m is the mass of the particle and $\langle v_\alpha^2 \rangle$ is the time averaged αth component of its velocity. This equivalence is often used to measure the temperature of a microcanonical system. Given that the temperature of a system is proportional (though not directly) to the average kinetic energy of the particles, the temperature can be controlled by scaling the velocities. If the temperature at time t is $T(t)$ and the velocities are multiplied by a factor λ, then the system temperature can be calculated by:

$$\Delta T = \frac{1}{2}\sum_{i=1} 2\frac{m_i(\lambda v_i)^2}{N_{df}k_B} - \frac{1}{2}\sum_{i=1} 2\frac{m_i v_i^2}{N_{df}k_B} \tag{2.13}$$

$$\Delta T = (\lambda^2 - 1)T(t) \tag{2.14}$$

$$\lambda = \sqrt{\frac{T_0}{T(t)}} \tag{2.15}$$

Multiplying the velocities at each time step by a factor $\lambda = \sqrt{T_0/T(t)}$, where $T(t)$ is the current temperature calculated from the kinetic energy and T_0 is the desired temperature provides the simplest way of keeping a constant system temperature.

A simpler formulation of velocity scaling is given by the Berendsen temperature coupling algorithm, in which the velocities are scaled at each step such that the rate of change of the temperature is proportional to the difference in temperature

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{2.16}$$

where τ is a parameter which determines how tightly the system is coupled to the thermal bath. The Berendsen temperature coupling algorithm gives an exponential decay of $T(t)$ towards

T_0. The coupling parameter τ is given by

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \cdot \left[\frac{T_0}{T(t - \frac{\delta t}{2})} - 1 \right] \tag{2.17}$$

In practice the coupling parameter τ is empirical and the choice of its value alters the strength of coupling between the system and the thermal bath and should be chosen within a reasonable range. Despite the empirical nature of the Berendsen coupling algorithm, it is very efficient for converging systems towards a desired temperature and consequently is commonly used in the equilibration step of an MD simulation[134]. Nevertheless, to accurately probe the canonical ensemble, the Extended System approach was first proposed simultaneously by Nosé[135] and Hoover[136]. The premise of the Nosé-Hoover approach is to an extended Lagrangian to consider the thermal bath as part of the system by adding a dynamic variable $\tilde{s}$, which has a non-zero mass and a velocity $\dot{\tilde{s}}$. In the extended system the atomic coordinates are identical to the non-coupled system, however the time scale is stretched by the factor $\tilde{s}$ so that

$$dt = \tilde{s}dt \tag{2.18}$$

The Lagrangian for the extended system is given as

$$L_{Nose} = \sum_{i=1}^{N} \frac{m_i}{2} s^2 \dot{r}_i^2 - U(r^2) + \frac{Q}{2}\dot{s}^2 - \frac{L}{\beta}\, ln\, \tilde{s} \tag{2.19}$$

where Q is the mass of $\tilde{s}$. The first two terms of the extended Lagrangian represent the potential energy subtracted from the kinetic energy of the real system. The third and forth terms represent the kinetic energy minus the potential energy assigned to $\tilde{s}$. The energy of the real system will fluctuate about a mean and accompanying the fluctuations of $\tilde{s}$, heat transfers occur between the system and the thermal bath, which regulate the system temperature. The momenta conjugate to $\mathbf{r}_i$ and s follow directly from Equ. 2.19:

$$\mathbf{p}_i \equiv \frac{\partial L}{\partial \dot{\mathbf{r}}_i} = m_i s^2 \dot{\mathbf{r}}_i \tag{2.20}$$

$$p_s \equiv \frac{\partial L}{\partial \dot{s}} = Q\dot{s} \tag{2.21}$$

This gives the extended Hamiltonian of the system if N particles plus the additional coordinate s:

$$H_{Nose} = \sum_{i=1}^{N} \frac{\dot{\mathbf{p}}_i^2}{2m_i s^2} + U\left(\mathbf{r}^2\right) + \frac{p_s^2}{2Q} + L\frac{ln\, s}{\beta} \tag{2.22}$$

Constant pressure (NPT)

Many experiments are conducted at a constant pressure, and rather than adjusting the volume of a simulation box in the canonical ensemble, it is often more convenient to ensure a constant pressure. Constant pressure simulations can be attained in the isothermal-isobaric ensemble by considering the volume as a dynamic variable that changes during the simulation. Similar to Hoover scheme to regulate the temperature of the water box, the volume parameter is accounted for by an extended Hamiltonian. The equations of motion for the positions and the momenta are[137]:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W}\mathbf{r}_i \tag{2.23}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \left(1 + \frac{d}{dN}\right)\frac{p_\epsilon}{W}\mathbf{p}_i - \frac{p_{\xi 1}}{Q_1}\mathbf{p}_i \tag{2.24}$$

where N is the number of particles; the thermostat variable, its momentum and mass are given by ξ_1, p_{ξ_1} and Q_1, respectively. This is similar to the thermostat in the Nosé-Hoover chain algorithm. Additionally, a barostat is introduced via the variables ϵ, p_ϵ and W, which give the additional variable, along with its momentum and mass, respectively. The dynamic

variable ϵ that accounts for the barostat is defined as the logarithm of the volume V of the system:

$$\epsilon = ln\left(\frac{V}{V_0}\right) \tag{2.25}$$

where V_0 is the volume of the system at time $t = 0$. The corresponding equations of motion for the volume are given by:

$$\dot{V} = \frac{dVp_\epsilon}{W} \tag{2.26}$$

$$\dot{p}_\epsilon = dV\left(P_{int} - P_{ext}\right) + \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} - \frac{p_{\xi 1}}{Q_1}p_\epsilon \tag{2.27}$$

where P_{ext} is the external pressure, which is selected at the outset of the calculation. The internal system pressure of the system (P_{int}) is given by:

$$P_{int} = \frac{1}{dV}\left[\sum_{i=1}^{N}\left(\frac{\mathbf{p}_i^2}{m_i} + \mathbf{r}_i \cdot \mathbf{F}_i\right) - dV\frac{\partial U(V)}{\partial V}\right] \tag{2.28}$$

The Hamiltonian of the extended system is given by:

$$H_{NPT} = H(\mathbf{p}, \mathbf{r}) + \frac{p_\epsilon^2}{W} + \sum_{i=1}\frac{p_{\xi 1}^2}{Q} + (dN+1)k_BT\,\xi_1 + k_BT\sum_{i=1}\xi_i + P_{ext}V \tag{2.29}$$

2.1.4 The functional form of the potential energy

When calculating the forces on each particle i using Equ. 2.7, it is necessary to evaluate the potential energy of each interaction in the system. The potential energy between each constitutive particle in a system is represented by an empirically-derived force field potential, which describes the interaction energy between atoms in terms of the atom coordinates ($\mathbf{r}$) and the force field parameters (s). The potential energy is written as a sum over different contribu-

tions, which can correspond to physical atomic interactions or to unphysical interactions that may be applied to the system:

$$U(\mathbf{r};s) = U^{phys}(\mathbf{r};s) + U^{special}(\mathbf{r};s) \tag{2.30}$$

Special unphysical interactions can correspond to restrains to bond lengths and distances, or additional forces applied to the system, though these will not be described here in any further detail. The physical interactions within the system can be further sub-divided into bonded and non-bonded interactions between the composite atoms:

$$U^{phys}(\mathbf{r};s) = U^{bonded}(\mathbf{r};s) + U^{non-bonded}(\mathbf{r};s) \tag{2.31}$$

The functional form of the bonded interactions of the Gromos force field parameter set, used herein, are the sum of the bond, bond angle, harmonic dihedral angle and trigonometric dihedral angle terms[138]:

$$U^{bonded}(\mathbf{r};s) = U^{bond}(\mathbf{r};s) + U^{bondangle}(\mathbf{r};s) + U^{harmonic}(\mathbf{r};s) + U^{trig.}(\mathbf{r};s) \tag{2.32}$$

The non-bonded interactions are a sum over the potential associated with van der Waals and electrostatic interactions between all pairs of atoms[138]:

$$U^{non-bonded}(\mathbf{r};s) = U^{vdW}(\mathbf{r};s) + U^{electrostatic}(\mathbf{r};s) \tag{2.33}$$

Bonded interaction potential

The energetic potential of covalent bonded interactions is calculated as the sum over all N_b bonds and depends on the parameters K_b and b_0, which were originally parametrised against experimental spectroscopic and X-ray diffraction data for small molecules. The functional

form of the bonded potential is:

$$U^{bond}(\mathbf{r}; K_b, b_0) = \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} \left[b_n - b_{0_n}\right]^2 \tag{2.34}$$

where b_n is the actual bond length, b_0 is the optimal bond length and K_b is the bond force constant. The parameters K_b and b_0 vary depending on the nature of the bonded interaction between given pairs of atoms (**Fig. 2.1 A**). The potential due to the angle between bonded interactions is calculated as a sum of all N_θ bond angles and depends on the parameters K_θ and θ_0:

$$U^{bondangle}(\mathbf{r}; K_\theta, \theta_0) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{b_\theta} \left[cos\, \theta_n - cos\, \theta_{0_n}\right]^2 \tag{2.35}$$

where θ_n is the actual bond angle, θ_0 is the optimal bond angle and K_θ is the bond angle force constant (**Fig. 2.1 B**). The harmonic torsion angles are used to maintain a specific planar configuration for four specific atoms (eg. maintaining a tetrahedral configuration about an sp3 hybridised carbon atom). This potential is calculated as the sum over all improper dihedral interaction centres N_ξ, with parameters K_ξ and ${\xi_0}^{138}$:

$$U^{harmonic}(\mathbf{r}; K_\xi, \xi_0) = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{b_\xi} \left[\xi_n - \xi_{0_n}\right]^2 \tag{2.36}$$

where ξ_n is the actual harmonic torsion (**Fig. 2.1 C**). Similarly, the trigonometric torsion angles are defined as a sum over all N_φ torsion angles, with parameters K_φ, δ and m:

$$U^{trig.}(\mathbf{r}; K_\varphi, \delta, m) = \sum_{n=1}^{N_\varphi} K_{\varphi_n} \left[1 + cos(\delta_n)\, cos(m_n \varphi_n)\right] \tag{2.37}$$

where δ is the phase shift, m_n is the multiplicity of the torsion angle and φ is the actual trigonometric torsion angle (**Fig. 2.1 D**).

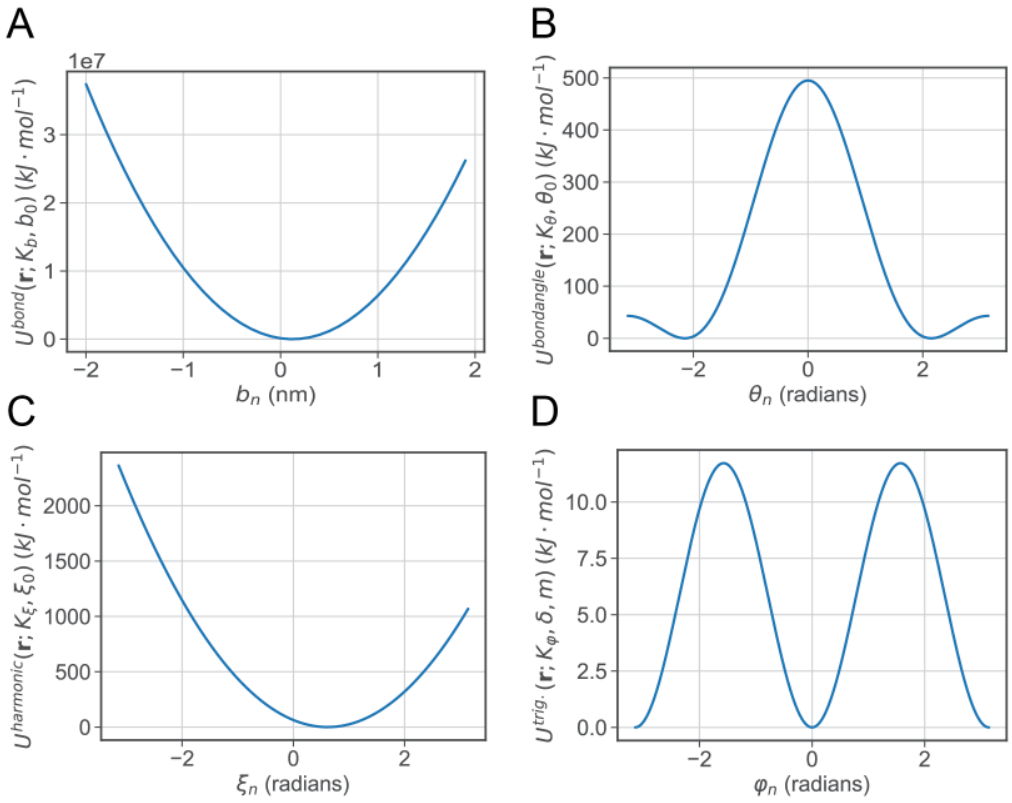

Figure 2.1: ***Examples of bonded interaction potential energies for arbitrarily selected atoms.*** *(A) The interaction potential for a C-O bond length. (B) The interaction potential for a H-N-C bond angle. (C) The harmonic torsional potential energy for a tetrahedral centre. (D) The trigonometric torsional potential energy for a -C-C- torsion.*

Non-bonded interactions

In the GROMOS force field parameter set, the non-bonded interactions are calculated as the potential between all atoms that are not involved in a covalent (ie. bonded) interaction. There are a number of exceptions to this collection of interactions, including those that fall outside of a defined inter-atomic cut-off range. Two principle non-bonded interaction types are considered: van der Waals interactions and electrostatic interactions. The non-bonded van der Waals potentials are calculated as a sum over all interacting non-bonded atom pairs i, j using a Lennard-Jones 12/6 interaction function with parameters $C12$ and $C6$:

$$U^{vdW}(\mathbf{r}; C12, C6) = \sum_{i,j} \left(\frac{C12_{i,j}}{r_{i,j}^{12}} - \frac{C6_{i,j}}{r_{i,j}^{6}} \right) \tag{2.38}$$

Electrostatic interaction potentials are defined as three separate terms: a Coulomb interaction potential, a distance-dependent reaction-field contribution and a distance-independent reaction-field term. The Coulomb potential is given as a sum over all interacting pairs U^C, with parameters q defined as the partial charges q_i on the atoms[138]:

$$U^C(\mathbf{r}; q) = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \cdot \frac{1}{r_{ij}} \tag{2.39}$$

where the dielectric permittivity of a vacuum and that of the solvent in which the atoms are embedded is given by ϵ_0 and ϵ_1, respectively. In addition to the direct Coulomb interactions, a reaction-field U^{RF} is determined, which represents the interaction of atom i with an induced dielectric field outside a defined cut-off distance R_{RF} in the presence of atom j. Finally, a distance-independent reaction-field term is given as a constant contribution to the potential for every pair of atoms taken into account.

Evaluating long-range electrostatic interactions

Increasing computational power has allowed for the simulation of ever larger molecular systems. Nevertheless, it is crucial that the computation of all pair interactions is avoided, as otherwise the time complexity of a molecular dynamics algorithm would be proportional to the square of the number of particles. The expense of a molecular dynamics simulation could be reduced rather drastically by truncating the potential of long-range electrostatics, thus assuming that the long-range part of the potential is unimportant. Truncation of long-range electrostatics removes the expensive part of the calculation but introduces serious inaccuracies. Other techniques are available for handling long-range electrostatic interactions and, while more expensive than a simple truncation, have the advantage that they more faithfully respect the long-range character of the forces between a charge-charge pair of atoms. Of the techniques available to evaluate long-range electrostatics, the so-called particle-mesh Ewald summation is often used for biomolecular systems and scales with a time complexity of $\mathcal{O}(n \log n)$.

The particle-mesh Ewald scheme is used as a numeric approximation to Ewald summation. In the Ewald summation method, long-range electrostatic interactions are split into two parts: a short-range contribution and a long-range contribution. The short-range contribution is solved in real-space, whereas the long-range contribution is calculated using a Fourier transform. The time complexity for a fully optimised Ewald summation scheme scales to the number of particles as $\mathcal{O}(n^{2/3})$, due to the expense of the reciprocal-space part of the Ewald sum. For the simulation of systems where the number of particles is of the order of 10^5 it is necessary to have a scheme that handles the Fourier part of the Ewald summation more efficiently. To this end, particle mesh Ewald methods allow for the efficient evaluation of the Fourier part of the Ewald summation scheme by using a so-called Fast Fourier Transform[139]. First, charged particles are assigned to discrete points on a lattice. Next, a Fast Fourier Transform method is used to compute the Poisson equation for the discrete charge distribution on the defined

lattice. Once Poisson equation has been computed to give the electrostatic energy, the forces are calculated and assigned back to the particles on the system.

2.2 Parameters used for molecular dynamics simulations of PKM2

Molecular dynamics (MD) simulations were used throughout this study to characterise the functional dynamics of PKM2 in explicit solvent and *in vacuo*.

2.2.1 Molecular dynamics simulations in explicit solvent

MD simulations of monomeric and tetrameric human PKM2 were performed with the GROMACS 5.2 engine[140], in SPC-E water[141]. Interactions between protein atoms and solvent were modelled using the Gromos 53a6 force field parameter set[138]. The input coordinates for monomeric PKM2 in the *apo* form (mPKM2^{apo}) were extracted from the Protein Data Bank crystal structure structure 3bjt[23]. Coordinates for mPKM2^{Phe} were extracted from 4fxj[22]; mPKM2^{FBP}, tPKM2^{FBP} and mPKM2$^{Tepp-46}$ from 3u2z[17]; and coordinates for tPKM2$^{FBP+Phe}$ and tPKM2$^{FBP+Ser}$ were extracted from 4b2d[19]. Missing residues were modelled using holomogy modelling within the Modeller suite[142]. The force-field parameters for FBP and Tepp-46 were determined using a quantum mechanical assignment of the partial charges using the ATB server[143]. Structures were solvated in a dodecahedral period box, such that the distance between any protein atom and the periodic boundary was a minimum of 1.0 nm. The system charge was neutralised by adding counter ions to the solvent (Na^+ and Cl^-). Equations of motion were integrated using the leap-frog algorithm with a 2 fs time step. The system was equilibrated for 5 ns in the NVT ensemble at 300 K and 1 bar. This was followed by a further 5 ns equilibration in the NPT ensemble. Following equilibration, multiple replicate production run simulations were performed for 400 ns under constant pressure and temperature conditions, 1 bar and 300 K. Temperature was regulated using the velocity-rescaling algorithm, with a coupling constant of 0.1. Covalent bonds and water molecules were restrained with the LINCS[144] and SETTLE[145] methods, respectively. Electrostatics were calculated with the

particle mesh Ewald method[139], with a 1.4 nm cut-off, a 0.12 nm FFT grid spacing and a four-order interpolation polynomial for the reciprocal space sums.

2.2.2 Molecular dynamics simulations *in vacuo*

Structural models of the 15+ monomer, 23+ A-A' and C-C' dimers and 33+ tetramers were generated from the PDB crystal structure 3bjt[23] by randomly assigning positive charges to histidine residues distributed throughout the protein. Models were simulated in vacuo using the OPLS- AA/L force-field parameter set[146]. Systems were minimised using the Steepest Descent algorithm for 5 x 106 steps, with a step size of 1 J mol^{-1} nm^{-1} and a maximal force tolerance of 100 kJ mol^{-1} nm^{-1}. Next, systems were equilibrated at consecutively increasing temperatures (100 K, 200 K and 300 K) each for 5 ns, with the Berendsen temperature coupling method and an integration step size of 1 fs. Production run simulations were performed for 10 ns with an integration step size of 2 fs in the canonical ensemble. Pressure coupling and electrostatics were turned off. Temperature was held constant at 300 K using the Berendsen coupling method. The most prevalent structures were extracted using the GROMOS clustering algorithm[147]. Theoretical collision cross sections were calculated for each clustered structure, using the projection approximation method, as outlined by Ruotolo (2008) *et al.*[148], and using the exact hard-sphere scattering model, as implemented in the EHSSrot software[149].

2.3 Protein biophysics

2.3.1 Recombinant protein expression and purification

PKM2 single-point mutant plasmids were generated through a single-step PCR reaction using hot-start KOD polymerase (Merck Millipore; Burlington MA, USA) and a pET28a-His-PKM2(WT) template plasmid (plasmid no. 42515 AddGene; Cambridge MA, USA). Plasmids were sequence-verified by Sanger Sequencing (Source Bioscience; Nottingham, UK). Expression of plasmids was achieved by transforming 40 ng pET28a-His-PKM2 into 40 μL *E. coli* BL21(DE3)pLysS (60413; Lucigen, Middleton WI, USA). Colonies were inoculated in LB media at 37 °C and grown to an optical density of 0.8 AU (600 nm). After bacteria had reached exponential growth, expression of the N-terminal His6-PKM2(WT) was induced with 0.5 mM isopropyl b-D-1 thiolgalactopyranoside (Sigma Aldrich, St. Louis MS, USA) at 24 °C for between 16 and 18 hours. The pellet was harvested and re-suspended in a lysis buffer consisting of 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 200 mM NaCl, 100 KCl and 10 mM imidizole, with the EDTA-free Complete protease inhibitor cocktail (Sigma Aldrich, St. Louis MS, USA). Cells were lysed by sonication at 4 °C. DNase was added at 1 μL/mL prior to centrifugation of the lysate at 20000 xg for 1 hour at 4 °C. The supernatant (the water-soluble cell fraction) was loaded onto a HisTrap HP nickel-charged IMAC column (GE; Boston MA, USA) and was washed with five column-volumes of wash buffer [10 mM HEPES pH 7.5, 10 mM $MgCl_2$, 100 mM KCl, 10 mM imidazole and 0.5 mM tris-2-carboxyethyl phosphine hydrochloride (TCEP; Sigma Aldrich, St. Louis MS, USA)]. After consecutive wash steps, the protein was eluted from the IMAC column with elution buffer buffer (10 mM HEPES pH 7.5, 10 mM $MgCl_2$, 100 mM KCl, 250 mM imidazole and 0.5 mM TCEP). The N-terminal His_6-epipope tag was cleaved with at 4 °C for 18 hours in cleavage buffer (50 mM Tris-HCl pH 8.0, 10 mM $CaCl_2$) with recombinant bovine thrombin, immobilised on agarose beads. Purified recombinant PKM2 was eluted from the thrombin-agarose column. Affinity purification was followed by size-exclusion chromatography on a HiLoad 16/60 Superdex 200 pg column (28-9893-35; GE, Boston MA, USA) at 500 mL/min flow rate with protein storage buffer (10

mM HEPES pH 7.5, 10 mM MgCl2, 100 mM KCl, and 0.5 mM TCEP) at 4 °C. Eluted PKM2 was collected and concentrated to a final protein concentration of 7 mg/mL with centrifugal concentrating filters (Vivaspin 20, 10 kDa molecular-weight cut-off, 28-9323-60; GE, Boston MA, USA). Protein purity was assessed by SDS-PAGE. The final concentration of the protein was obtained by measuring the fluorescence absorbance spectrum between 240 nm and 450 nm. The concentration was estimated (molar extinction coefficient of 29.910 M^{-1} cm^{-1} at 280 nm).

†Oligo I.D	Sequence
JM.G122P.F	CCTGAGATCCGAACTCCGCTCATCAAGGGCAGC
JM.I124G.F	ATCCGAACTGGGCTCGGCAAGGGCAGCGGCACT
JM.G204P.F	AATGGTGGCTCCTTGCCGAGCAAGAAGGGTGTG
JM.G204A.F	AATGGTGGCTCCTTGGCGAGCAAGAAGGGTGTG
JM.F244V.F	ATGGTGTTTGCGTCAGTGATCCGCAAGGCATCT
JM.R246Q.F	TTTGCGTCATTCATCCAGAAGGCATCTGATGTC
JM.R246A.F	TTTGCGTCATTCATCGCGAAGGCATCTGATGTC
JM.K247P.F	GCGTCATTCATCCGCCCGGCATCTGATGTCCAT
JM.D288R.F	ATCCTGGAGGCCAGTGCGGGGATCATGGTGGCT
JM.D288N.F	ATCCTGGAGGCCAGTAACGGGATCATGGTGGCT
JM.K305Q.F	GAGATTCCTGCAGAGCAGGTCTTCCTTGCTCAG
JM.F307P.F	CCTGCAGAGAAGGTCCCGCTTGCTCAGAAGATG
JM.F307A.F	CCTGCAGAGAAGGTCGCGCTTGCTCAGAAGATG
JM.C326S.F	GGGAAGCCTGTCATCAGCGCTACTCAGATGCTG
JM.A327S.F	AAGCCTGTCATCTGTAGCACTCAGATGCTGGAG
JM.A327D.F	AAGCCTGTCATCTGTGATACTCAGATGCTGGAG
JM.D357S.F	GTCCTGGATGGAGCCAGCTGCATCATGCTGTCT
JM.C358A.F	CTGGATGGAGCCGACGCGATCATGCTGTCTGGA
JM.G435A.F	GTCCTCACCAAGTCTGCGAGGTCTGCTCACCAG
JM.G435P.F	GTCCTCACCAAGTCTCCGAGGTCTGCTCACCAG
JM.R489Q.F	GAGGACGTGGACCTCCAGGTGAACTTTGCCATG
JM.R489L.F	GAGGACGTGGACCTCCTGGTGAACTTTGCCATG
JM.F492A.F	GACCTCCGGGTGAACGCGGCCATGAATGTTGGC

†Identifier: JM.<single point mutant substitution>.<forward (F) / reverse (R)>

Table 2.1: Forward primer sequences for single-point mutants of PKM2

†Oligo I.D	Sequence
JM.G122P.R	GCTGCCCTTGATGAGCGGAGTTCGGATCTCAGG
JM.I124G.R	AGTGCCGCTGCCCTTGCCGAGCCCAGTTCGGAT
JM.G204P.R	CACACCCTTCTTGCTCGGCAAGGAGCCACCATT
JM.G204A.R	CACACCCTTCTTGCTCGCCAAGGAGCCACCATT
JM.F244V.R	AGATGCCTTGCGGATCACTGACGCAAACACCAT
JM.R246Q.R	GACATCAGATGCCTTCTGGATGAATGACGCAAA
JM.R246A.R	GACATCAGATGCCTTCGCGATGAATGACGCAAA
JM.K247P.R	ATGGACATCAGATGCCGGGCGGATGAATGACGC
JM.D288R.R	AGCCACCATGATCCCCGCACTGGCCTCCAGGAT
JM.D288N.R	AGCCACCATGATCCCGTTACTGGCCTCCAGGAT
JM.K305Q.R	CTGAGCAAGGAAGACCTGCTCTGCAGGAATCTC
JM.F307P.R	CATCTTCTGAGCAAGCGGGACCTTCTCTGCAGG
JM.F307A.R	CATCTTCTGAGCAAGCGCGACCTTCTCTGCAGG
JM.C326S.R	CAGCATCTGAGTAGCGCTGATGACAGGCTTCCC
JM.A327S.R	CTCCAGCATCTGAGTGCTACAGATGACAGGCTT
JM.A327D.R	CTCCAGCATCTGAGTATCACAGATGACAGGCTT
JM.D357S.R	AGACAGCATGATGCAGCTGGCTCCATCCAGGAC
JM.C358A.R	TCCAGACAGCATGATCGCGTCGGCTCCATCCAG
JM.G435A.R	CTGGTGAGCAGACCTCGCAGACTTGGTGAGGAC
JM.G435P.R	CTGGTGAGCAGACCTCGGAGACTTGGTGAGGAC
JM.R489Q.R	CATGGCAAAGTTCACCTGGAGGTCCACGTCCTC
JM.R489L.R	CATGGCAAAGTTCACCAGGAGGTCCACGTCCTC
JM.F492A.R	GCCAACATTCATGGCCGCGTTCACCCGGAGGTC

†Identifier: JM.<single point mutant substitution>.<forward (F) / reverse (R)>

Table 2.2: Reverse primer sequences for single-point mutants of PKM2

2.3.2 Spectrophotometric assay to determine molar amounts of fructose 1,6-bisphosphate

Purified recombinant PKM2 was heat-precipitated at 90 °C and supernatant was carefully extracted. The supernatant was subsequently analysed for amounts of fructose 1,6-bisphosphate (FBP) using an assay with three coupled enzymatic steps. The reaction mixture contained 20 mM Tris-HCl pH 7.0, 150 μM NADH, 0.7 U·mL^{-1} glycerol 3-phosphate dehydrogenase (G-3-PDH), 7 U·mL^{-1} triose phosphate isomerase (TPI) and the supernatant of 5-50 μM purified recombinant PKM2 after heat-precipitation at 90 °C. Reactions were initiated by adding between 0.008 and 0.016 U·mL^{-1} to a total reaction volume of 100 μL. The enzymes rabbit muscle TPI and G-3-PDH were purchased as a mixture from Sigma (50017, Sigma Aldrich; St. Louis MS, USA). Rabbit muscle aldolase was also purchased from Sigma (A2714, Sigma Aldrich; St. Louis MS, USA). For each molecule of FBP consumed in the reaction, two molecules of NADH are oxidised. Therefore, FBP amounts were quantified by monitoring NADH oxidation over time at 25 °C in a 1 mL quartz cuvette (1 cm path-length) by measuring the NADH absorption signal at 340 nm using a Jasco V-550 UV-Vis spectrophotometer. For the assay calibration, known amounts of FBP from a powder stock were used instead of the heat-precipitated PKM2 supernatant.

2.3.3 Measurement of PKM2 steady-state enzyme kinetics

Initial velocities for the forward reaction of pyruvate kinase [phosphoenolpyruvate (PEP) and adenosine diphosphate (ADP) conversion to pyruvate and adenosine triphosphate (ATP) (**Fig. 2.2**)] were measured using a reaction coupled to rabbit muscle lactate dehydrogenase. The reaction monitored the oxidation of NADH (ϵ_{340nm} = 6220 M^{-1} cm^{-1}) at 37 °C in a buffer containing 10 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$ and 0.5 mM TCEP. Initial velocity versus substrate concentrations for PEP were measured in the absence and in the presence of allosteric ligands, in a reaction buffer containing 180 μM NADH and 8 U·mL^{-1} rabbit muscle lactate dehydrogenase (Sigma; St. Louis MS, USA). Reactions were initiated by adding PEP at the desired concentration, with ADP at a constant concentration of 5 mM.

A total protein concentration of 5 nM PKM2 was used for all enzyme reactions, in a total volume of 100 μL per well in a 96-well plate. Apparent kinetic constants K_M and k_{cat} were determined by fitting initial velocity curves to Michaelis-Menten steady-state kinetic models.

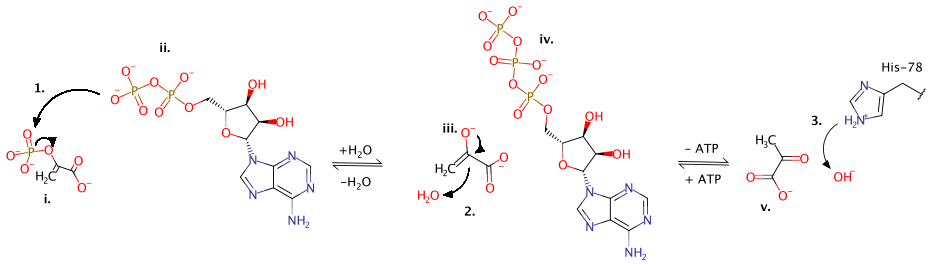

Figure 2.2: ***The catalytic mechanism of the pyruvate kinase-catalyzed conversion of phosphoenolpyruvate and ADP to pyruvate and ATP. 1.*** *A phosphoryl group is transferred from phosphoenolpyruvate (i) to ADP (ii) by an apparent S_{N_2} mechanism, to yield the enolate of pyruvate (iii) and ATP (iv).* ***2.*** *A water molecule at the active site protonates the enolate (iii) at the 2-si face of the double bond to form keto pyruvate (v.).* ***3.*** *the hydroxide is reprotonated by the active site residue His-78. Divalent cations Mg^{2+} or Mn^{2+} are thought to enhance the acidity of the solvent molecule[150]. K^+ does not directly contact the substrate or intermediate, but instead is thought to influence the structure of the active site through interactions with R72, R119 and K269[55].*

2.3.4 Analysis of the steady-state kinetics of PKM2 enzyme activity inhibition by Phe

The dependence of the PKM2 enzyme kinetic constants K_M, k_{cat} and $\frac{k_{cat}}{K_M}$ on the concentration of Phe were determined, in order to assign the mechanism by which Phe inhibition of PKM2 occurs in the absence and in the presence of FBP. Steady-state measurements of PKM2 enzyme activity (as described in Section 2.3.3) were performed by titrating the substrate phosphoenolpyruvate at several different concentrations of phenylalanine and a constant concentration of 5 mM adenosine diphosphate. In order to investigate the allosteric K-type effect of Phe on enzyme affinity for its substrate PEP, a single-substrate-single-effector paradigm was assumed. Under prevailing equilibrium conditions the rate equation of the general modifier mechanism reveals apparent values of K_M and k_{cat}:

$$\frac{v}{[E]_t} = \frac{k_{cat}^{app}[S]}{K_M^{app} + [S]} \tag{2.40}$$

$$\frac{v}{[E]_t} = \frac{k_2 \frac{1+\beta\frac{[X]}{\alpha K_X}}{1+\frac{[X]}{\alpha K_X}}[S]}{K_M \frac{1+\frac{[X]}{K_X}}{1+\frac{[X]}{\alpha K_X}+[S]}} \tag{2.41}$$

where $[E]_t$ is the concentration of enzyme active sites, X is the inhibitor (Phe), S is the substrate, K_X is the dissociation constant of the specific component of the enzyme mechanism, α is the reciprocal allosteric coupling constant and β is the factor by which the inhibitor affects the catalytic constant k_2. It then follows that the dependence of the equilibrium constants K_M^{app}, k_{cat}^{app} and $(\frac{k_{cat}}{K_M})^{app}$ on the concentration of a modifier (X) (the allosteric inhibitor phenylalanine) are written as follows[151]:

$$k_{cat}^{app} = k_2 \cdot \frac{1+\beta\frac{[X]}{\alpha K_X}}{1+\frac{[X]}{\alpha K_X}} \tag{2.42}$$

$$K_M^{app} = K_M \cdot \frac{1 + \frac{[X]}{K_X}}{1 + \frac{[X]}{\alpha K_X}} \tag{2.43}$$

$$\left(\frac{k_{cat}}{K_M}\right)^{app} = \frac{k_2}{K_M} \cdot \frac{1 + \beta\frac{[X]}{\alpha K_X}}{1 + \frac{[X]}{K_X}} \tag{2.44}$$

Rate curves measuring the dependence of [Phe] on the three kinetic constants were fit to the equations above to solve for K_s, K_x, α and β. The kinetic mechanism was assigned based on the topology of rate-modifier mechanisms detailed by Baici (2015)[151].

2.3.5 Measurement of the allosteric coupling co-efficient

PKM2 initial velocities were measured at 37 °C using a lactate dehydrogenase assay over a range of phosphoenolpyruvate concentrations with a constant concentration of 5 mM ADP, as previously described in Section 2.3.3. Activity measurements were repeated following pre-incubation of the PKM2 variant with saturating concentrations of FBP (2 μM for all variants, with the exception of R489L which was incubated with 50 mM FBP to saturate this variant). The allosteric coupling constant (Q) was calculated to determine the coupling between FBP binding and catalysis, as previously described[152]:

$$Q = \frac{K_{ia}}{K_{ia/x}} \tag{2.45}$$

where K_{ia} and $K_{ia/x}$ are equilibrium dissociation constants for the binding of the substrate (a) in the absence or presence, respectively, of the allosteric effector (x). Q > 1, indicates positive allosteric coupling between the binding of x to the protein and the binding of a to the substrate binding pocket. Conversely, where Q < 1, negative coupling exists between the a and x sites. Measurements were repeated after addition of 400 μM Phe to the protein variants that had been pre-incubated with FBP, and activity was measured over a range of substrate

concentrations.

2.3.6 Circular dichroism spectroscopy

A JASCO J-815 spectrometer was used to record far-UV circular dichroism (CD) spectra (Jasco; Oklahoma City, OK USA) from 200 nm to 260 nm with 300 μL of 0.2 mg/mL PKM2 ion a quartz cuvette with a path length of 0.1, at a temperature of 20 °C. Raw data in units of mdeg were converted to extinction co-efficient of the mean residue CD extinction co-efficient, in units of M^{-1} cm^{-1}:

$$\Delta\epsilon_{mrw} = \frac{S \cdot MRW}{32980 \cdot c \cdot L} \tag{2.46}$$

where c is the molar concentration of the protein, L is the path length, S is the raw measurement of CD intensity (in units of mdeg) and MRW is the molecular weight of the protein divided by the number of amino acids in the protein.

To measure the thermal stability of recombinant protein, the CD intensity at 222 nm was monitored over a range of temperatures. Melting curves were fit to a two-state equilibrium model, which assumes that the protein is either folded or unfolded and that there is no intermediate (semi-folded) state ever populated:

$$F \underset{K_f}{\overset{K_u}{\rightleftharpoons}} U \tag{2.47}$$

where N is the native state, U is the unfolded state, K_u is the unfolding constant and K_n is the folding constant. The unfolding constant is defined as:

$$K_u = \frac{[F]}{[U]} = \frac{F_f}{F_u} = \frac{1 - F_f}{F_f} \tag{2.48}$$

The CD absorbance signal at 222 nm over a range of temperatures, starting from the folded and ending in the unfolded state, is assumed to reflect a linear combination of detected optical

signals from the folded and the unfolded species in solution:

$$S_{obs} = S_f F_f + S_u F_u = \frac{S_f + S_u K_u}{1 + K_u} = \frac{S_f + S_u e^{\frac{-\Delta G}{RT}}}{1 + e^{\frac{-\Delta G}{RT}}} \tag{2.49}$$

where S is the signal and F is the fractional content of either species. Substituting for ΔG with the Gibbs-Helmholtz equation gives:

$$S_{obs} = \frac{S_f + S_u exp\left[-\frac{\Delta H_{T_m}}{RT}\left(1 - \frac{T}{T_m}\right) - \frac{\Delta C_p}{RT}\left((T - T_m) - T \cdot ln\left(\frac{T}{T_m}\right)\right)\right]}{1 + exp\left[-\frac{\Delta H_{T_m}}{RT}\left(1 - \frac{T}{T_m}\right) - \frac{\Delta C_p}{RT}\left((T - T_m) - T \cdot ln\left(\frac{T}{T_m}\right)\right)\right]} \tag{2.50}$$

where T_m is the melting temperature, ΔH_{T_m} is the enthalpy change at $T = T_m$ and ΔC_p is the heat capacity change.

2.3.7 Measurements of FBP binding to PKM2

PKM2 binding to FBP was measured by titrating a concentrated solution stock solution of FBP into 5 μM PKM2 in a buffer containing 10 mM HEPES pH 7.0, 100 mM KCl and 10 mM $MgCl_2$ at 20 °C. Intrinsic fluorescence emission spectra of PKM2 were recorded using a Jasco FP-8500 fluorescence spectrofluorometer with an excitation wavelength of 280 nm (bandwidth of 2 nm) and emission scanned from 290 nm to 450 nm (bandwidth of 5 nm) in a 0.3 cm path length quartz cuvette (Hellma Analytics; Muellheim, Germany). A ratio of the emission intensities at 325 and 350 nm was plotted against the concentration of the titrant. Binding curves were fit to a model assuming a 1:1 binding stoichiometry (1 FBP molecule per monomer of PKM2) with a non-linear least squares regression fit of the following:

$$S_{obs} = S_P P_0 + S_L L_0 + (S_{PL} - S_P - S_L) \cdot \frac{(K_D + P_0 + L_0) - \sqrt{(K_d + P_0 + L_0)^2 - 4(P_0 L_0)}}{2} \tag{2.51}$$

where the spectral signal S_{obs} is the ratio of fluorescence emissions at 325 nm and 350 nm; S_P, S_L, and S_{PL} are the spectral signals of the free protein, the free ligand and the protein-ligand complex, respectively. The apparent dissociation constant is given by K_D. The free protein

concentration ($[P_0]$) was corrected by subtracting the percentage of protein pre-bound to co-purified FBP, as determined from the aldolase-coupled assay.

To automate the solution of Equ. 2.51 for binding data, an R package *ligBind* was written. The package consists of two simple functions 1. to fit binding data to Equ. 2.51 and 2. to plot the binding curve along with the raw data and the residuals of the fit to Equ. 2.51. A brief description of ligBind is given here, which will provide a tutorial for how to use the package to estimate the binding affinity for a ligand-receptor interaction. *ligBind* can be compiled from source and is freely available to download at `https://github.com/jamieAmacpherson/ligand_binding/tree/master/ligBind`. Once *ligBind* has been compiled, the user can fit binding data to estimate the affinity and visualise the binding curve, in an automated manner. For the purposes here we will use test data that is shipped with the package.

After the package has been compiled, the user can calculate the binding affinity of the protein-receptor interaction using the function fit.binding(). The fitting function requires as inputs: (1) a data-frame with two equal columns containing the ligand concentration and the binding response, (2) a predicted binding affinity and (3) the protein concentration used in the experiment. The function fit.binding() prints the calculated binding affinity along with the associated error. The function also returns a four-element list containing the summary of the fitting procedure, the raw binding data, the fitted binding model and the residuals between the fitted model and the raw data. This list can be used as a direct input for the plotting function ligBind::plt.binding.curve() to visualise the binding data and the predicted model. An example output of the plotting function in *ligBind* is shown in **Fig. 2.3**.

Listing 1: Use of *ligBind* to calculate binding affinity and visualise ligand-receptor binding data.

```
## Load ligBind package
library('ligBind');

## Initialise the test data
test_data = dat;

## Calculate the binding affinity of the protein-receptor
## interaction.
test_data_fit = ligBind::fit.binding(bindingdat = test_data,
                kd_pred = 3000,
                prot_conc = 5);

## Visualise the binding curve
ligBind::plt.binding.curve(test_data_fit);
```

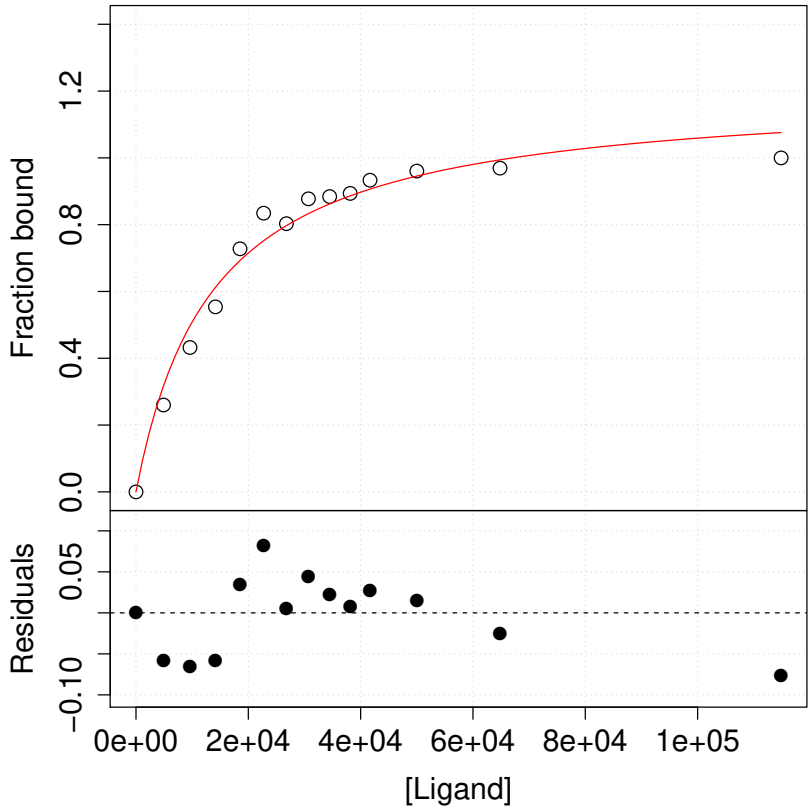

Figure 2.3: ***Visualisation of a ligand-receptor binding curve fitted using the ligBind package.*** *The output of ligBind::plt.binding.curve() function shown above in Listing 1.*

2.3.8 Microscale thermophoresis measurements of phenylalanine and serine binding

Microscale thermophoresis (MST) is an experimental technique that is used to measure protein-ligand binding. In this technique, a protein-ligand mixture is prepared in a buffered solution and is loaded into a glass capillary. During the MST experiment, a temperature gradient is induced by an infrared laser, which induces protein-ligand molecules to disperse from the focal point of the laser. The directed movement of the protein-ligand molecules is detected by monitoring the fluorescence absorption or emission properties at a defined wavelength using either covalently attached fluorophores or the intrinsic fluoresence of either the protein or the ligand. The thermophoretic properties of a protein is strongly dependent on its size, charge, hydration shell and conformation. Therefore, protein-ligand binding which perturbs any of these biomolecular properties can be detected using MST.

The binding of Phe and Ser to PKM2 was measured using MST on a Monolith NT.115 instrument (Nanotemper Technologies; Munich, Germany). PKM2 was fluorescently labelled with an Atto-647 fluorescein dye (NT-647-NHS; Nanotemper, Munich, Germany). 250 μL of 20 μM recombinant PKM2 was labelled with 250 μL of 60 μM dye in a buffer containing 100 mM bicarbonate pH 8.5 and 50 % DMSO for 30 minutes at room temperature in the dark. Free dye was separated from labelled PKM2 using a NAP-5 20 ST size-exclusion column (GE; Boston MA, USA). The stoichiometry of labelled- to unlabeled-PKM2 was determined by measuring the fluorescence absorbance ratios at 647 nm (labelled; PKM2-NT495NHS) and 280 nm (unlabeled PKM2). Thermophoresis measurements of labelled PKM2 were obtained using a Monolith NT.115 instrument (Nanotemper; Munich, Germany). Labelled PKM2, at a constant concentration of 30 nM, was titrated with either Phe (up to 5 mM) or Ser (up to 10 mM) in a buffer containing 10 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.5 mM TCEP and 0.1 % tween-20. Prior to each thermophoresis measurement, capillary scans were obtained to determine sample homogeneity. Apparent binding curves were fit assuming a 1:1 stoichiometry.

2.4 Native mass spectrometry

Micro Bio-Spin 6 chromatography colums (Bio-Rad Laboratories, Hercules, CA, US) were used to buffer-exchange PKM2 samples into 200 mM ammonium acetate (Fisher Scientific, Loughborough, UK). Protein samples were diluted to a concentration of 5 μM - 20 μM. Ligands were dissolved in 200 mM ammonium acetate and added to the protein prior to MS analysis.

Three separate instruments were used for native mass spectrometry experiments: an Ultima Global (Micromass, UK) extended for high mass range, a modified Synapt G2 (Waters; Wilmslow, UK) where the triwave assembly was replaced with a linear drift tube and a Synapt G2-Si. Positive ionisation mode was used in analysing samples. Nano-electrospray ionisation was applied from borosilicate glass capillary tips, pulled in-house on a Flaming/Brown P-1000 micropipette puller (Sutter Instrument Company, Novato CA, USA). A platinum wire was inserted inside the tip after the protein solution had been loaded, to allow the application of a positive voltage. All voltages were kept as low as possible to achieve spray while keeping the protein in a native-like state. Typical conditions used were a capillary voltage of $\simeq$ 1.2 kV, a cone voltage of $\simeq$ 10 V and a source temperature of 40 °C.

2.4.1 Mass-deconvolution of native spectra

The mass-deconvolved spectrum o was estimated from protein spectra y by maximising the conditional probability $P(o|y)$ using the Bayes Rule:

$$P(o|y) = P(y|o)P(o) \tag{2.52}$$

The prior probability of the mass spectrum $P(o)$ is given by:

$$P(o) = \frac{N!}{o_1!, o_2!, ..., o_k!} Z_1^{o_1} Z_2^{o_2}, ..., Z_k^{o_k} \tag{2.53}$$

where Z is defined by an initial guess of the parent mass spectrum. The probability for the native m/z spectrum y given the mass spectrum o is given by the following term:

$$P(y|o) = \prod_i exp\left[-\frac{(y_i - a_i)^2}{\sigma^2}\right] \tag{2.54}$$

2.4.2 Ion mobility mass spectrometry

An in-house modified Synapt G2 was used for IM-MS measurements, in which the original triwave assembly was replaced with a linear drift tube with a length of 25.05 cm. Drift times were measured in a helium buffer gas at a temperature of 298.15 K and a pressure of 1.99 - 2.00 torr. Motilities for all charge states were converted into rotationally averaged collision cross sections ($^{DT}CCS_{He}$) using the Mason-Schamp equation:

$$K = \frac{3q}{16N} \cdot \left(\frac{1}{m} + \frac{1}{M}\right)^{\frac{1}{2}} \left(\frac{2\pi}{k_B T}\right)^{\frac{1}{2}} \frac{1}{\Omega} \tag{2.55}$$

where K is the measured mobility, q is the charge of the analyte ion, N is the density of the buffer gas, m is the mass of the analyte ion, M is the mass of the buffer gas, k_B is the Boltzmann constant, T is temperature and Ω is the rotationally-averaged collision cross section. The collision cross sections of each charge-state species were further converted into a single collision cross section distribution in which all charge states contribute towards the overall distribution, in proportion to their intensity in the mass spectrum[153].

2.5 Measurement of metabolite concentrations in human cell lines

2.5.1 Liquid chromatography-mass spectrometry (LC-MS) detection of metabolites

Samples were injected into a Dionex UltiMate liquid chromatography system (Thermo Scientific; Waltham MA, USA) with a pHILIC (150 mm x 4.6 mm, 5 μm particle size) column (Merk Sequant; Millipore Sigma, Burlington MA, USA). An elution gradient was used with a 80:20 solvent mixture composed of 20 mM ammonium carbonate (solvent A) and acetonitrile (solvent B). The elution gradient was run over 15 minutes, followed by a 5 minute wash with a solvent mixture of 95:5 solvent A to solvent B. Additional parameters: 10 μL injection volume; auto-sampler temperature 4 °C; flow rate 300 μL min^{-1}; column temperature 25 °C. Positive/negative polarity switching was performed using a Q Exactive Orbitrap (Thermo Scientific; Waltham MA, USA) with a HESI II (Heated electrospray ionisation) probe. The following mass spectrometry parameters were used: spray voltage 3.5 kV (positive mode) and 3.2 kV (negative mode); probe temperature 320 °C; sheath and auxiliary gases 30 and 5 arbitrary units, respectively. A full scan range between 70 m/z to 1050 m/z with settings of AGC target (3×10^6) and a mass resolution setting of 'Balanced and High (70,000). The Xcalibur 3.0.63 software suite (Thermo Scientific; Waltham MA, USA) was used for data acquisition. Prior to data analysis, a Thermo Scientific Calmix solution was used as a standard to perform mass calibration in both electrospray ionisation polarities and ubiquitous low-mass contaminants were used to apply lock-mass correction to each analytical run in order to enhance calibration stability. Parallel reaction monitoring (PRM) acquisition parameters: resolution 17,500 (ion count), auto gain control target 2×10^5, maximum isolation time 100 ms, isolation window m/z 0.4. The collision energies were set individually in the high-energy collisional dissociation mode. In order to assess the stability and performance of the system, quality controls were performed by extracting an equal volume of each sample and pooling them. This quality control mixture was subsequently analysed throughout the run. Data analysis

was performed using the Xcalibur Qual Browser and Tracefinder 4.1 software suites (Thermo Scientific; Waltham MA, USA).

2.5.2 Metabolite extraction and cell volume calculations

Cells were seeded in 6 cm dishes in RPMI media containing 10 % dialysed foetal calf serum, 24 hours prior to the start of the experiment. An hour prior to treatment of cells, the media was refreshed and then changed again at the start of the experiment to RPMI with or without 11 mM glucose; or to Hank's Balanced Salt Solution (HBSS) with or without supplemented amino acids as described in the text. Four technical replicate plates were used for each condition, and 2-4 plates for each cell line were used to count cells and measure mean cell diameter. This was then used to determine the intracellular volume of the cells, and subsequently the intracellular concentrations of the detected metabolites. Media treatments were performed for 1 hour, after which plates were washed twice with ice-cold 120 mM phosphate buffer saline (PBS), and 725 μL of ice-cold methanol was used to quench the cells. The cells were scraped from the plates and were transferred into 180 μL H_2O and 160 μL $CHCl_3$. A further 725 μL methanol was used in a second scraping of each plate. The final methanol-chloroform-water mixtures containing the cells were vortexed and sonicated in a sonicating water bath at 4 °C. Extraction of the metabolites was allowed to proceed at 4 °C for 16 hours, before sedimenting precipitated material and then drying down the supernatant. Polar and apolar phases were separated by re-suspending the dried metabolites in a 1:3:3 mixture of chloroform-to-methanol-to-water, with a total volume of 350 μL. Polar metabolites were then analysed by LC-MS (as detailed above). For absolute quantification of the metabolites of interest, known quantities of ^{13}C-labelled versions of those metabolites were spiked into lysates. Amounts of the unlabelled compounds could then be determined as a proportion of the intensity given by the labelled compound. Previously determined cell numbers and volumes were then used to determine the intracellular concentrations of each of the metabolites of interest.

Metabolite	Formula	Mass (Da)	Pos. mode m/z	Neg. mode m/z	Mode used	†R.T (min)
FBP	$C_6H_{14}O_{12}P_2$	339.99611	341.00394	338.9829	neg.	13.42
$^{13}C_6$-FBP	$^{13}C_6H_{14}O_{12}P_2$	346.01621	347.02404	345.00839	neg.	13.42
PEP	$C_3H_5O_6P$	167.98241	168.99023	166.97458	neg.	13.58
$^{13}C_6$-PEP	$^{13}C_2C_1H_5O_6P$	169.98911	170.99693	168.98128	neg.	13.58
Ser	$C_3H_7NO_3$	105.0426	106.05043	104.03478	pos.	13.14
$^{13}C_6$-Ser	$^{13}C_3H_7NO_3$	108.05265	109.06048	107.04483	pos.	13.14
Phe	$C_9H_{11}NO_2$	165.07899	166.08681	164.07116	pos.	9.7
$^{13}C_6$-Phe	$^{13}C_3C_6H_{11}NO_2$	171.09909	172.10691	170.09126	pos.	9.7

† R.T is an abbreviation for the retention time.

Table 2.3: Liquid-chromatography mass spectrometry analysis parameters used for the detection of ^{12}C and ^{13}C metabolite standards.

2.6 Absolute quantitation of intracellular PKM2 amounts using targeted proteomics

The absolute amounts of trypsin-digested PKM1 and PKM2 peptides, isolated from four cancer cell lines (MCF7, LN229, SN12C and HCT116), were measured using a targeted proteomics approach. The cell lines used for these experiments have previously been shown to highly express PKM2. The exon-9/10 splicing event of the *pkm* transcript is mutually exclusive and so either PKM1 or PKM2 is expressed by a given cell. Nevertheless, to control for amounts of both PKM1 and PKM2 protein in the cells, proteotypic peptides for both isoforms were synthesised, along with several peptides that mapped to both the PKM1 and PKM2 primary sequences. A pilot experiment identified five peptides to be used for the quantitation of PKM2/1. These peptide standards were synthesised by the Crick Proteomics STP with arginine and lysine labelling (Arginine: $^{13}C_6$, $^{15}N_4$; and Lysine: $^{13}C_6$, $^{15}N_2$).

Peptide	Sequence	Specificity
1	GDLGIEIPAEK	PKM1/PKM2
2	APIIAVTR	PKM1/PKM2
3	ITLDNAYMEK	PKM1/PKM2
4	LAPITSDPTEATAVGAVEASFK	PKM2
5	LFEELVR	PKM1

All peptides were synthesised with arginine and lysine heavy labelling (Arginine: $^{13}C_6$, $^{15}N_4$; and Lysine: $^{13}C_6$, $^{15}N_2$).

Table 2.4: Proteotypic peptide standards used for absolute quantification of PKM2 in cell lysates.

2.6.1 Trypsin digestion

Cell lysates containing a total of 50 μg of protein content were precipitated by adding ice-cold acetone. Samples were centrifuged at 8000 xg for 10 minutes at 4 °C, following overnight incubation at -20 °C. The supernatant was removed and any residual acetone was evaporated at room temperature. The pellet was dissolved in 50 mM TEAB, reduced with 10 mM DTT and alkylated with 20 mM iodoacetamide. Next, protein was digested with 1 μg of trypsin overnight at 37 °C, after which each sample was spiked with a mixture of the five heavy-labelled peptide standards (Table 2.4). For mass spectrometry analysis, 1 μg of peptide was loaded onto a 50 cm Easy Spray C18 column (Thermo Scientific; Waltham MA, USA).

2.6.2 Analysis of peptides by LC-tandem MS (LC-MS/MS

A Dionex U3000 system (SRD3400 degasser, WPS-3000TPL-RS autosampler, 3500RS nano pump) coupled to a QExactive electrospray ionisation hybrid quadruole-orbitrap mass spectrometer (Thermo Scientific; Waltham MA, USA) was used for mass spectrometic analysis. Reverse-phase chromatography with a buffer of 5 % DMSO in 0.1 % formic acid (mobile phase A) and 5 % DMSO, 80 % acetonitrile in 0.1 % formic acid (mobile phase B) was used at a flow rate of 250 nL·min^{-1}. Digested samples were run of linear gradient of mobile phase B in 90 minutes, giving a total run time (including column conditioning) of 120 minutes. The nanoLC was coupled to a QExactive mass spectrometer using an EasySpray nano source (Thermo Scientific; Waltham MA, USA), with a spray voltage of +2.1 kV, a capilary temperature of 250 °C an an S-lens radio-frequency level of 55 AU.

Chapter 3

Enzyme kinetics and ligand binding biophysical studies reveal a combined role for multiple allosteric ligands in the regulation of PKM2.

3.1 Introduction

PKM2 enzyme activity is regulated by a number of intracellular metabolites. The upstream glycolytic intermediate fructose 1,6-bisphosphate (FBP) stimulates PKM2 catalysis by binding to an allosteric pocket in the C-domain of the protein[55,65,70,71]. Additionally, serine (Ser) and histidine (His) increase; whereas phenylalanine (Phe), alanine (Ala), tryptophan (Trp), valine (Val) and proline (Pro) inhibit PKM2 enzyme activity by binding to a common pocket, sandwiched between the A- and the C-domains[19,20,22,64,65]. Crystallographic structures of PKM2 showing that the FBP- and amino acid binding pockets are physically and chemically distinct, suggests that PKM2 can concurrently bind to both FBP and an amino acid[19]. What remains unclear, however, is how simultaneous binding of multiple ligands with, opposing functional effects, control PKM2 enzyme activity.

The following chapter investigates the functional consequences of allosteric regulation by a number of regulators, both alone and in combination, on PKM2 enzyme activity. In support of *in vitro* findings, we explore cellular metabolic conditions required for concurrent allosteric regulation of PKM2.

3.2 Characterising the kinetics of FBP binding to and activation of PKM2

Activation of PKM2 by FBP is a prototypical and long-studied example of feed-forward regulation in cellular metabolism[154]. Nevertheless, the mechanism of PKM2 activation is disputed; some studies have reported an increased affinity for the substrate phosphoenolpyruvate (PEP)[19,47–49] and others reporting changes to the rate of substrate turnover[20,22]. To reconcile the differing results in the literature, and to characterise PKM2 regulation at the basal level, we started by investigating the kinetics of FBP binding and regulation of PKM2 activity.

3.2.1 FBP binds to PKM2 with nanomolar affinity

PKM2 contains three tryptophan residues, two of which (W482 and W515) are proximal to the FBP binding pocket[55]. Previous studies have reported that addition of FBP leads to changes to the intrinsic fluorescence emission spectrum of PK[155], likely resulting from side-chain conformational changes to W482 and W515 upon FBP binding[22]. Therefore, FBP binding to PKM2 was measured by monitoring tryptophyl fluorescence (λ_{EX} = 280 nm, λ_{EM} = 350 nm).

Measurements of PKM2 fluorescence suggested sub-stoichiometric FBP binding

We found that titrating PKM2 with a stock solution of FBP resulted in a ligand concentration-dependent excitation and red-shifting of the protein emission spectrum (**Fig. 3.1 A**). The effect of FBP on the fluorescence of PKM2 was sufficiently large to allow for a determination of the dissociation constant from titrations of the fluorescence change. Nevertheless, an inspection of the fluorescence spectra revealed that full saturation of the PKM2 fluorescence response to FBP addition occured at ligand concentrations less than the concentration of protein used in the titration (**Fig. 3.1 A**), suggesting sub-stoichiometric binding despite a 1:1 stoichiometry reported in several crystal structures[17,19,55]. We therefore hypothesised that purification of recombinant PKM2 may co-purify residual amounts of FBP produced by the *E. coli* expression

vector, which had been previously reported[22,23,63].

Sub-stoichiometric amounts of FBP are co-purified with recombinant PKM2

Preparations of purified recombinant PKM2 were assessed for their content of co-purified FBP using a three-step aldolase assay[156], which couples the conversion of FBP into glyceraldehyde 3-phosphate (GA-3-P) and dihydroxyacetone phosphate (DHAP) with the oxidation of NADH by glycerol 3-phosphate dehydrogenase (G-3-PDH), requiring the conversion of GA-3-P to DHAP by triosephosphate isomerase (TPI) as an intermediate step (**Fig. 3.1 B**). The assay was calibrated by quantifying molar amounts of FBP from a known solution of the compound made up from powder (**Fig. 3.1 C**). Subsequent quantitation of FBP in several independent preparations of purified PKM2, following heat-precipitation of the protein, revealed that up to 77.5 % of purified PKM2 was pre-bound to FBP (**Fig. 3.1 D** and **E**), consistent with previous reports of FBP co-purification with PKM2[23,63]. Many attempts to purify recombinant PKM2 under denaturing conditions, and thereby remove co-purified ligand, were unsuccessful (data not shown). Therefore, all subsequent experiments were performed with preparations of PKM2 with less than 25 % of the protein pre-bound to FBP, so that experimental data could be interpreted consistently. Experiments performed on purified PKM2 in the absence of any added ligands, though containing quantified sub-stoichiometric amounts of FBP, are denoted with an asterisk (Apo$*$).

Determination of the free protein concentration was used to calculate the K_D^{FBP}

Quantification of the FBP-PKM2 saturation status using the aldolase assay allowed for an estimate of the concentration of free protein in the purified PKM2 sample, which was essential for accurately calculating the binding constant in the ligand titration experiment because the fluorescence signal is produced by the *free* protein P_0 upon association with the *free* ligand L_0 in solution. A variable proportion of the total PKM2 (P_{total}) was sequestered in a protein-ligand complex PL and would therefore not produce a fluorescence change upon further addition of

the ligand. Therefore, the free protein concentration of a given preparation of PKM2 could be calculated:

$$P_0 = P_{total} - PL \tag{3.1}$$

It was unclear, however, whether sub-stoichiometric pre-occupancy of FBP meant that ligand binding was distributed amongst the protomers, producing a mixed population of tetramers with varying FBP-binding stoichiometries (0:4, 1:4, 2:4, 3:4 and 4:4), or whether a fixed percentage of PKM2 tetramers were fully bound to FBP. The distribution of stoichiometries in the starting material was likely determined by the cooperativity of FBP binding to PKM2; in an infinitely cooperative system, binding of a limiting pool of ligand to a multi-site oligomeric protein would be expected to be all or none. Nevertheless, it was assumed that the binding affinities of FBP to each of the four structurally identical pockets in the PKM2 homo-tetramer was identical, and produced equivalent fluorescence spectral changes upon ligand binding. Moreover, it was posited that the PKM2 fluorescence change was a result of ligand binding only, and that any potential oligomeric changes would not contribute to the emission spectrum. Therefore, the changed PKM2 fluorescence signal, observed upon addition of FBP, is given by a mixture of fluorescence signals from P_0 and PL:

$$S_{obs} = S_P \cdot [P_0] + (S_{PL} - S_P) \cdot [PL] \tag{3.2}$$

where S_P and S_{PL} are optical properties of P_0 and PL, respectively. The concentration of PL at equilibrium is given by:

$$[PL] = \frac{(K_D + [P_0] + [L_0]) - \sqrt{(K_D + [P_0] + [L_0])^2 - 4([P_0][L_0])}}{2} \tag{3.3}$$

where K_D is the dissociation constant of the FBP-PKM2 interaction. Substituting the expression for $[PL]$ back into Equ. 3.2:

$$S_{obs} = S_P[P_0] + (S_{PL} - S_P) \cdot \frac{(K_D + [P_0] + [L_0]) - \sqrt{(K_D + [P_0] + [L_0])^2 - 4([P_0][L_0])}}{2} \quad (3.4)$$

After correcting for the partial pre-occupancy of FBP, a solution for the quadratic function $S_{obs}(S_P, S_L, S_{PL}, P_0, L_0, K_D)$ (Equ. 3.4) was calculated using a non-linear least squares procedure to provide an estimate of the dissociation constant K_D. The calculation of the binding affinity of a protein-ligand interaction was automated in an R package *ligBind* (see Methods section 2.3.7 for details and software tutorial). An average apparent K_D^{FBP} of (21.4 ± 9.0) nM was estimated from a total of six replicate FBP binding measurements (**Fig. 3.1 F**).

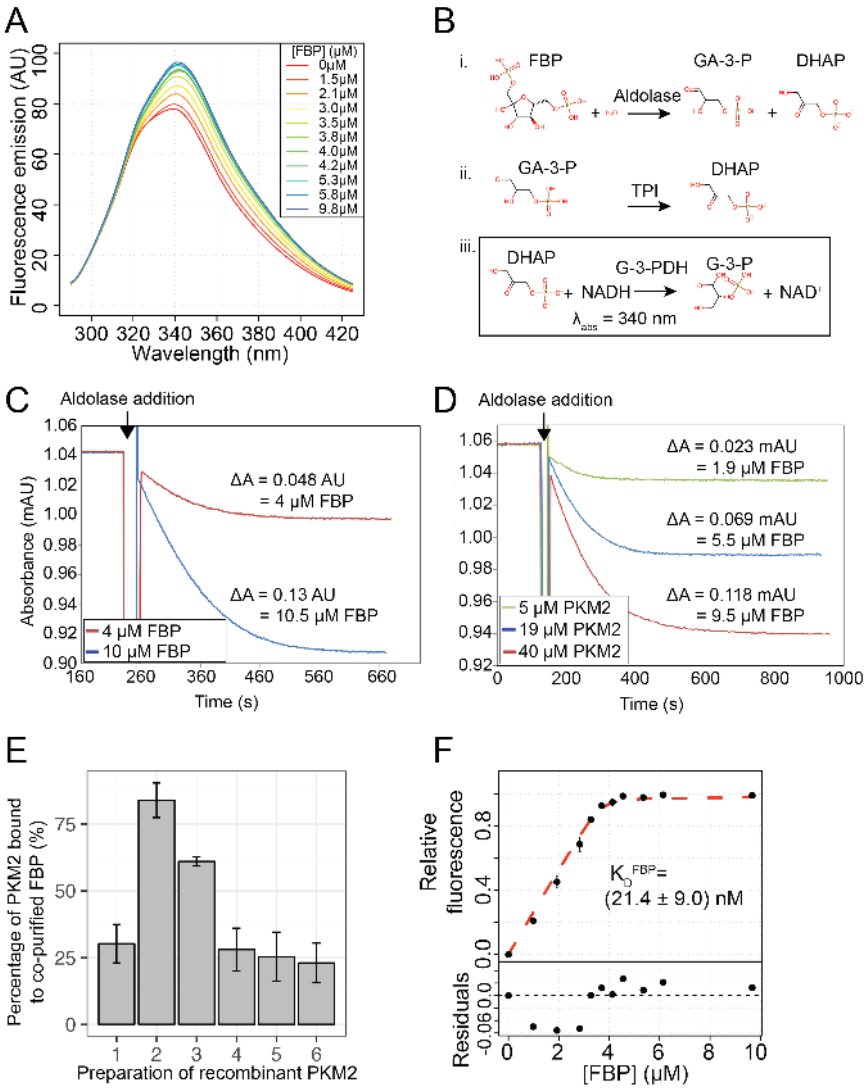

Figure 3.1: ***FBP binds to PKM2 with nano-molar affinity.*** ***(A)*** *Fluorescence spectra of human PKM2 in 10 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$ at 20 °C. A protein concentration of 5 μM was used for all curves and the concentration of FBP is specified in the legend.* ***(B)*** *Molar amounts of FBP co-purified with recombinant PKM2 were quantified using a three-step enzymatic assay, as described in the text.* ***(C)*** *Calibration of the aldolase assay to estimate the sensitivity for measuring amounts of co-purified FBP. The assay was tested by adding 4 μM or 10 μM purified FBP.* ***(D)*** *Preparations of 5 μM (green), 19 μM (blue) and 40 μM (red) of purified PKM2 were heat-precipitated at 90 °C to release any co-purified FBP. The supernatant was subsequently analysed for its molar contents of FBP, as previously described in (B).* ***(E)*** *Quantitation of the percentage saturation of PKM2 with FBP in six independent preparations of purified recombinant PKM2.* ***(F)*** *Binding was monitored from spectroscopic measurements of PKM2 fluorescence emission with increasing concentrations of FBP. The relative changes to emission at 325 nm and 350 nm is plotted against the concentration of added FBP. The binding affinity was calculated, as described in the text. Means and standard deviations from six separate experiments are plotted.*

3.2.2 Technical note: the protein concentration determines the error associated with the K_D of PKM2-FBP binding

Experimental measurements of PKM2 fluorescence in Section 3.2.1 were necessarily performed at a protein concentration of 5 μM because of the limited quantum-yield of tryptophan. As such, P_0 was approximately 220-times greater than the estimated K_D^{FBP}. Measurements of ligand binding are known to be increasingly imprecise under conditions where $[P_0] >> K_D$ because small additions of L are sequestered into the PL complex (i.e $L_0 << L_{total} \simeq PL$)[157,158], which was accounted for in our numerical treatment of the binding data (see Section 3.2.1). We therefore questioned to what extent using a protein concentration orders of magnitude in excess of the binding constant would contribute the imprecision of the affinity estimate.

To this end, a numerical approach was employed whereby theoretical ligand-protein binding data were simulated using the ligBind::simtitr() function within a new R package *ligBind* (see Methods section 2.3.7). Binding data were simulated over a range of increasing P_0 and a defined *theoretical* K_D of 1 μM. The simulated binding data were subsequently fit using Equ. 3.4 to estimate a *calculated* K_D, which was compared to the initialised *theoretical* K_D to determine the percent error of the affinity calculate for a given P_0. Where the theoretical $K_D >> P_0$, the transition between unbound and fully bound is achieved over a gradient of ligand concentrations (**Fig. 3.2 A**). In contrast, where $K_D << P_0$, the resulting quadratic function $S_{obs}(S_P, S_L, S_{PL}, PL, K_D)$ has a sharp apex, leading to a higher percent error (**Fig. 3.2 B**).

Next, the K_D was calculated for simulated binding data over a range of $[P_0]$ up to 250 μM. We found that, where $[P_0] \leq 50 \cdot K_D$ the calculated binding affinity was approximately equivalent to the true binding affinity when fitted with the quadratic function $S_{obs}(S_P, S_L, S_{PL}, PL, K_D)$ (**Fig. 3.2 C**), resulting in a negligible percent error of the apparent affinity. As the free protein concentration was increased such that $P_0 \geq 50 \cdot K_D$, we observed a non-linear increase in the calculated binding affinity (μK_D) (**Fig. 3.2 C**).

For comparison, the same binding data were fitted with a hyperbolic function of the form:

$$S_{obs} = \frac{[L_{total}]}{[L_{total}] + K_D} \tag{3.5}$$

similar to other classical fitting methods, where binding is expressed as a function of the total ligand concentration ($[L_{total}]$), but does not account for the concentration of the three species present at equilibrium (P_0, L_0 and PL). Fitting simulated binding data over a range of P_0 using a classic graphical approach in Equ. 3.5, high P_0 leads to an unacceptable percentage error of > 10000 % (**Fig. 3.2 D**), which dwarfs the relatively moderate error we obtain by fitting with the quadratic function in Equ. 3.4 (**Fig. 3.2 C**).

Therefore, given the experimental limitations (outlined above) that precluded us from using PKM2 concentrations close to the K_D, the fit of the quadratic equation permitted the use of our binding data to determine K_Ds. Importantly, **Fig. 3.2 C** suggests that at higher P_0, the calculated K_D^{FBP} reported in Section 3.2.1 is likely higher than the theoretical K_D (i.e. FBP binds even tighter than our data suggest).

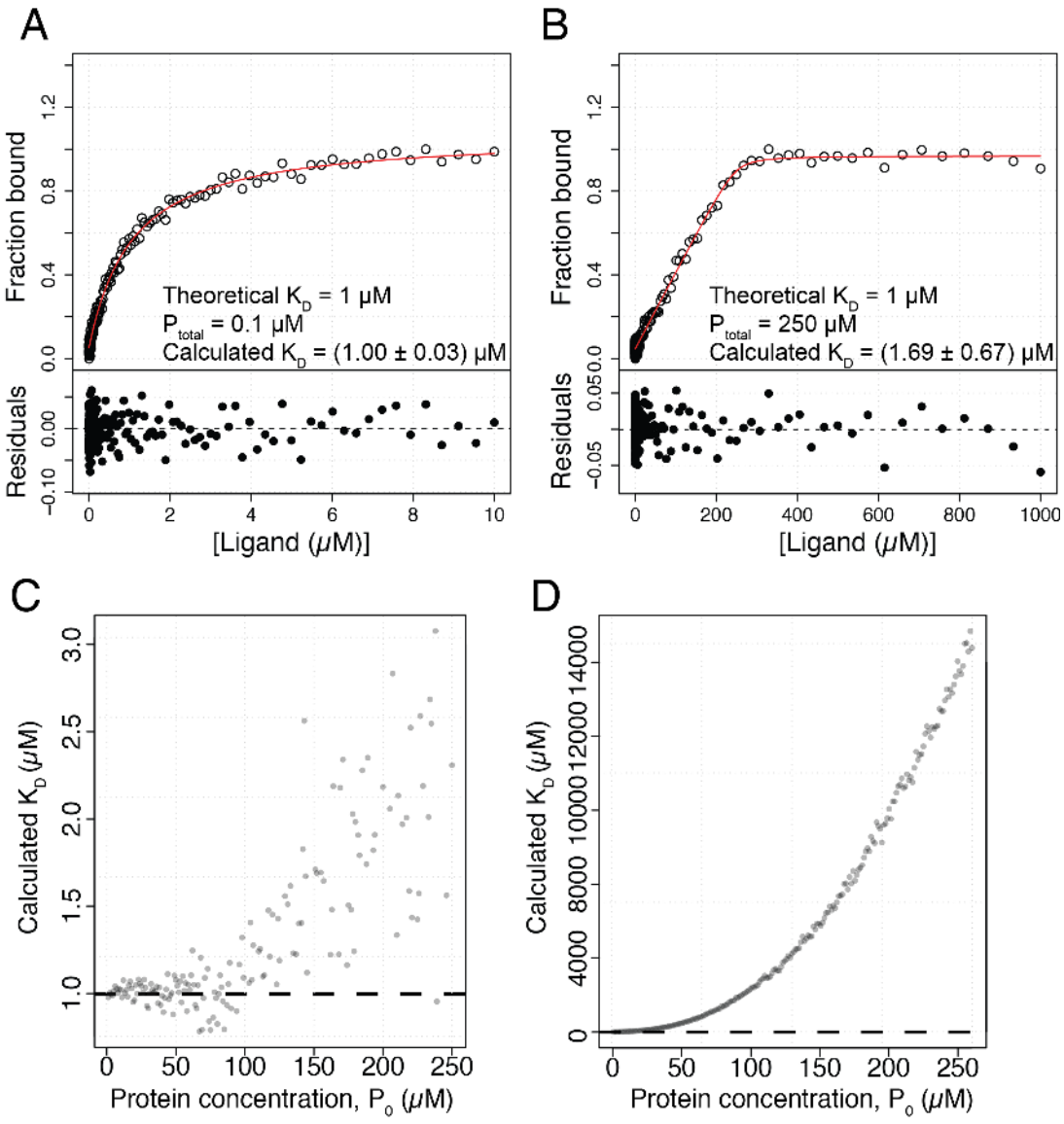

Figure 3.2: ***Simulated protein-ligand binding data show that the measured binding affinity is over-estimated when high concentrations of protein are used in the experiment.*** ***(A)*** *Binding data (open circles) were simulated with a defined theoretical K_D of 1 μM and a protein concentration (P_0) of 0.1 μM (i.e $K_D >> P_0$). A binding curve was calculated from the simulated data (red curve), using the quadratic function $S_{obs}(S_P, S_L, S_{PL}, PL, K_D)$ described in Equ. 3.4.* ***(B)*** *Binding data were simulated as in (A) with a defined theoretical K_D of 1 μM and a protein concentration (P_0) of 250 μM (i.e $K_D << P_0$).* ***(C)*** *Binding affinities were calculated for simulated binding data with a theoretical KD = 1 μM over a range of P_0 concentrations, using the quadratic function $S_{obs}(S_P, S_L, S_{PL}, PL, K_D)$ (Equ. 3.4) or* ***(D)*** *using the hyperbolic function $S_{obs}(K_D)$ (Equ. 3.5).*

3.2.3 FBP binding leads to an increase in the affinity of PKM2 for its substrate

To investigate the mechanism of FBP-induced activation, steady-state kinetic parameters of PKM2 were determined in the presence and absence of FBP. In the absence of any added FBP, the K_M^{PEP} was (1.22 ± 0.02) mM with a k_{cat} of (349.3 ± 40.9) s^{-1} (**Fig. 3.3 A** and **B**). Addition of saturating concentrations of FBP resulted in a six-fold reduction in the K_M^{PEP} to (0.23 ± 0.04) mM, without a significant change in k_{cat} (**Fig. 3.3 A** and **B**). The activation constant (AC_{50}^{FBP}) for FBP was measured as (118.1 ± 19.0) nM (**Fig. 3.3 C**), in excess of our previously measured K_D^{FBP}. Taken together, steady-state measurements of PKM2 catalysis suggested, consistent with previous findings, that FBP acts as a K-type activator to increase the binding affinity of the catalytic substrate PEP, without changing the rate of substrate turnover (k_{cat}).

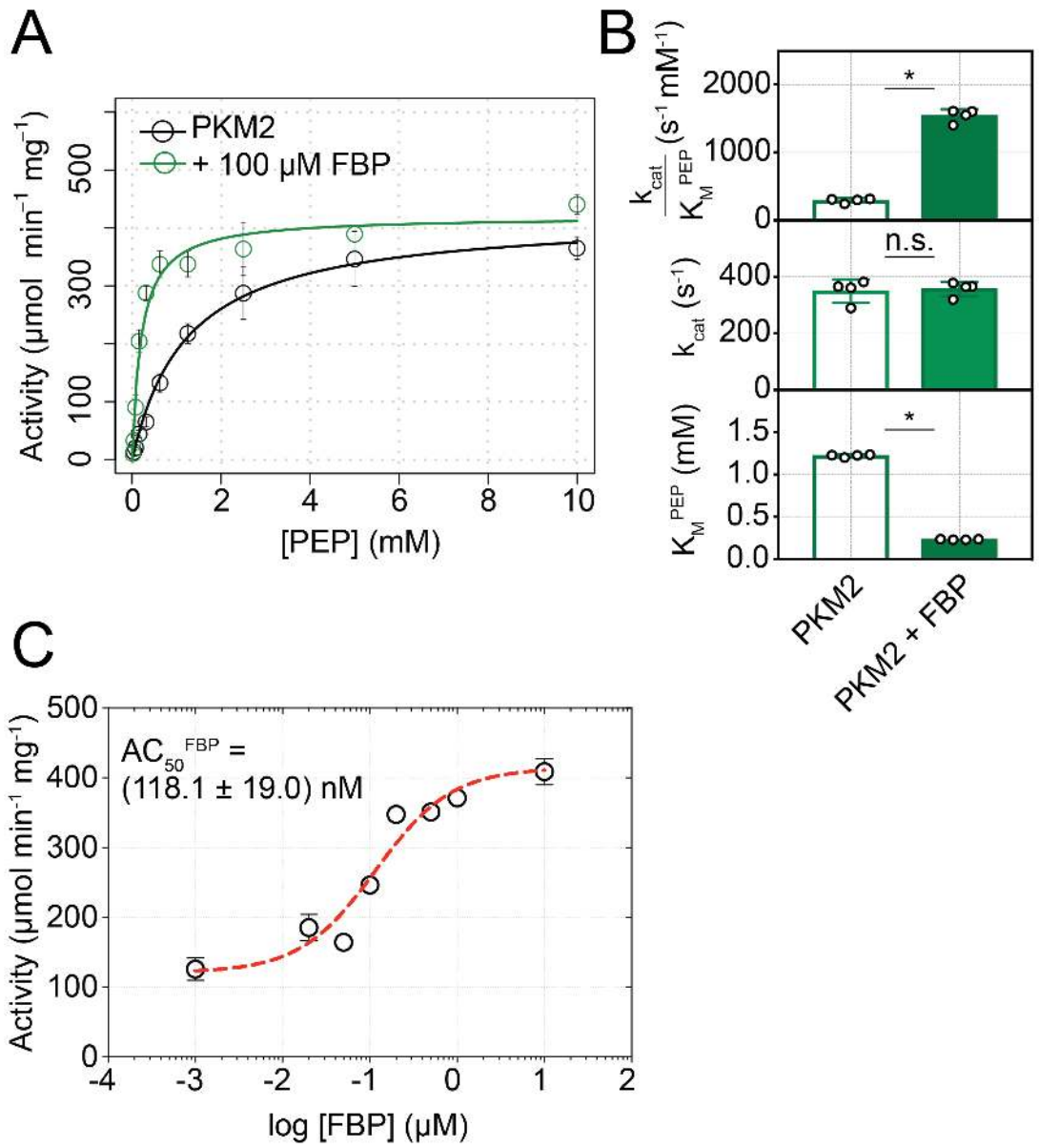

Figure 3.3: ***Addition of FBP results in an increase in the substrate affinity of PKM2 for phosphoenolpyruvate.*** ***(A)*** *Specific activity of 5 nM PKM2 was measured using an LDH-coupled spectrophotometric assay (see Methods Section 2.3.3) in the absence (black) and in the presence (green) of 2 μM added FBP. A constant concentration of 5 mM ADP was used, while varying the concentration of PEP. Measurements were performed at 37 °C. Rate curves were fitted using Michaelis-Menten kinetics. Means and standard deviations from four separate experiments are plotted.* ***(B)*** *Kinetic parameters were quantified from activity measurements of PKM2 in the absence (black) and in the presence (green) of 2 μM FBP. Significance was assessed using a Wilcoxon rank-sum test. Asterisk (*) marks significant changes (p-value < 0.05).* ***(C)*** *Specific activity of 5 nM PKM2 was measured at 1.5 mM PEP and 5 mM ADP, varying the concentration of FBP. A binding curve was fitted to the resulting rate curve assuming a 1:1 stoichiometry, to estimate the apparent activation constant (AC_{50}^{FBP}). Means and standard deviations of four separate experiments are plotted.*

3.2.4 A phosphotyrosine peptide binds competitively with FBP to PKM2

Given the finding that FBP binds to PKM2 with nano-molar affinity *in vitro* (Section 3.2.1), consistent with previously reported K_D^{FBP} measurements[59,62,63], we sought to investigate possible mechanisms by which the affinity is reduced by competition with other ligands. Growth factor signalling mediated by phosphorylation of proteins at tyrosine residues, which is often upregulated in cancer, inhibits PKM2 activity by outcompeting FBP with a phosphotyrosine peptide motif[23]. The mechanism of PKM2 phosphotyrosine peptide regulation, however, and the kinetics of its competition with FBP have not been characterised. We therefore sought to investigate the kinetics with which the phosphotyrosine peptide (M2tide) identified by Christofk *et al.* (2008)[23] displaces FBP.

M2tide binds to PKM2 with an apparent affinity of 150 μM

To this end, competition binding experiments between FBP and M2tide were performed by measuring the fluorescence anisotropy (λ_{EX} = 494 nm, λ_{EM} = 518 nm) of a fluorescein-labelled M2tide variant (fluor-M2tide)[159,160]. The sequence of fluor-M2tide is given: Fluorescein-GGAVDDD(PTyr)AQFANGG-COOH.

Preliminary experiments found that addition of PKM2 resulted in an increase in the anisotropy of fluor-M2tide (data not shown). Subsequent titrations of fluor-M2tide with PKM2 resulted in a concentration-dependent increase in the anisotropy of the peptide, which were used to calculate an apparent binding constant [$K_D^{fluor-M2tide}$ = (153 ± 19) μM] (**Fig. 3.4 A**). Measurements were repeated with a single-point mutant PKM2(K433E), previously shown to disrupt phosphotyrosine binding to PKM2[23]. Consistent with this, we found that fluor-M2tide did not bind to PKM2(K433E), though a linear increase in the fluorescence anisotropy of the fluorescein probe was observed, likely a non-specific concentration-dependent effect of the labelled peptide (**Fig. 3.4 B**). The notion that the linear fluorescence signal of the peptide arose from a non-specific effect, was supported by the observation that titrating the peptide

with purified bovine serum albumin (BSA) resulted in a similar linear fluorescence anisotropic signal (**Fig. 3.4 B**).

M2tide and FBP competitively bind to PKM2

To investigate the competitive binding kinetics between fluor-M2tide and FBP, fluorescence measurements of FBP-PKM2 binding were performed (as previously in Section 3.2.1) in the presence of increasing concentrations of fluor-M2tide. We found that addition of the peptide was accompanied by a linear increase the apparent K_D^{FBP} (**Fig. 3.4 C**), suggesting that the peptide and FBP competitively bind to PKM2. Conversely, while the K433E mutant abolished binding of fluor-M2tide, FBP binding was maintained albeit with a lower affinity [K_D = (6.46 ± 1.46) μM] (**Fig. 3.4 D**). Taken together, these results suggest that fluor-M2tide binds adjacent to the FBP pocket and that residue K433 provides a critical charged interaction for its binding, which would imply that the binding pockets for fluor-M2tide and FBP partially overlap.

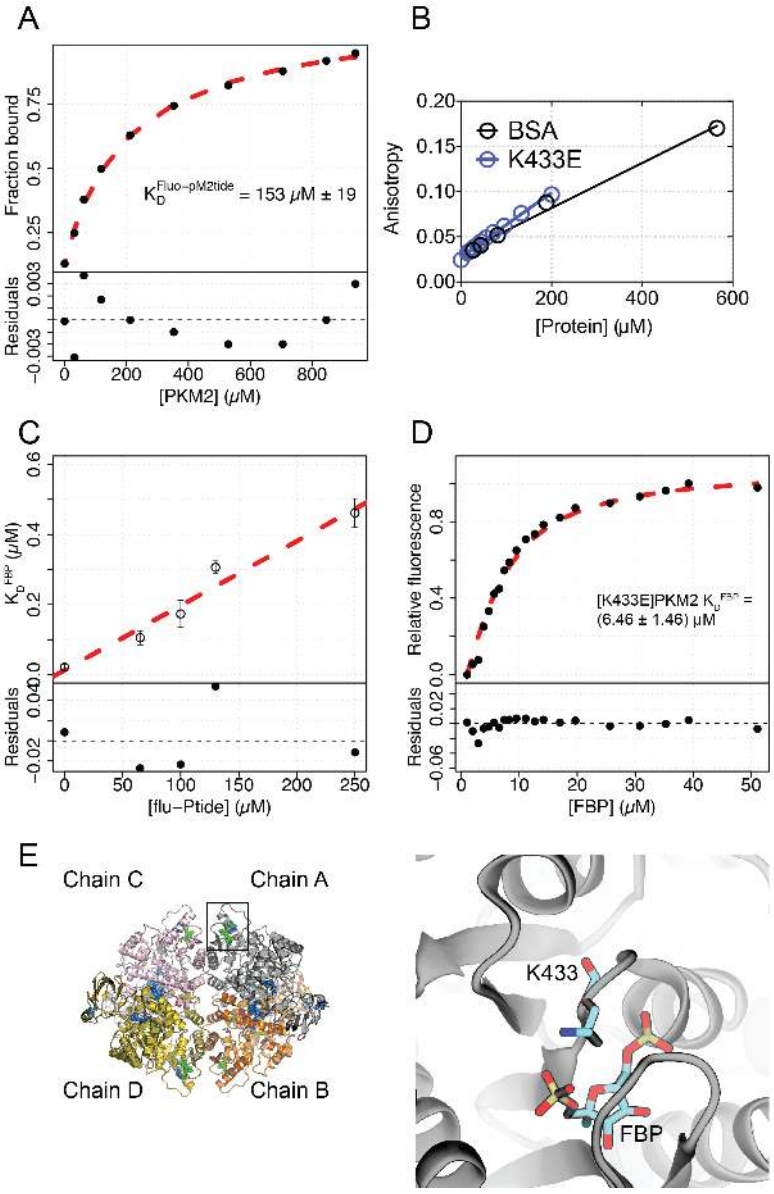

Figure 3.4: ***Fluor-M2tide competes with FBP for binding to PKM2.*** ***(A)*** *Fluorescence anisotropy measurements of a fluorescein-labelled peptide (fluor-M2tide) were performed at increasing concentrations of recombinant PKM2. The apparent binding affinity was estimated from a non-linear least squares regression of the binding curve.* ***(B)*** *Fluorescence anisotropy measurements of fluor-M2tide were acquired, titrating the concentrations of either PKM2(K433E) (blue) or bovine serum albumin (black). Anisotropy measurements were fitted with a linear regression.* ***(C)*** *Fluorescence emission measurements of PKM2 were performed to estimate the binding affinity of FBP at varying concentration of the fluor-M2tide peptide. Means and standard deviations from three independent titrations are plotted.* ***(D)*** *Fluorescence emission spectroscopy measurements of PKM2(K433E) were performed over a range of FBP concentrations.* ***(E)*** *A structural view of residue K433 relative to FBP bound in the crystal structure. Structural coordinates were extracted from the PDB accession code 3u2z.*

3.3 FBP binding to PKM2 induces inter-protomeric cooperativity

Consistent with previous reports of FBP-induced activation of PKM2[55,59,161], the previous Section found a K-type mechanism whereby FBP binding increases the effective affinity of the protein for the substrate phosphoenolpyruvate. The reciprocal relationship between FBP binding and enzyme activation, however, is further complicated by the reported association of monomers and/or dimers into stable tetramers, which may invoke ligand-induced communication pathways between the four active sites in the tetramer forming the basis of inter-protomeric cooperativity. Alternatively, enzyme catalysis and subsequent activation of PKM2 protomers may act in isolation within the purported tetramer assembly.

3.3.1 Titrating PKM2(WT) with the catalytically-dead PKM2(R72A) results in a non-linear decay in enzyme activity

A mutant-doping assay can be used to quantify inter-subunit cooperativity of a hetero-oligomeric protein

In order to test the hypothesis that FBP-induced allosteric activation promotes inter-subunit co-operativity, we sought to monitor the kinetics in the decay of PKM2 activity as PKM2(WT) was titrated with a catalytically-inactive mutant variant of PKM2, in a so-called *mutant doping* assay[162–165]. In this assay, PKM2(WT) was titrated with a catalytically-inactive mutant variant of PKM2 to form hetero-oligomeric species with defined stoichiometric ratios of WT subunits and inactive subunits. The enzyme activity of each mixture is measured, which decreases with increasing concentrations of the inactive species in the mixture. For proteins with a cooperative dependence on the adjacent subunits in an oligomeric assembly, the decrease in activity is non-linear. Conversely, for proteins where the subunits function independently of their neighbouring subunits within an oligomeric assembly, the activity decay is linear. Therefore, the inter-subunit of an oligomeric protein can be quantified by fitting the decay in enzyme activity as the WT protein is *doped* with an inactive variant.

PKM2(R72A) is inactive and hetero-oligomerises with PKM2(WT)

To first identify a single-point mutant that depleted PK phosphotransferase activity, the enzyme activity of three active site mutants [PKM2(R72A), PKM2(R120A) and PKM2(K270A)] were measured. Initial velocity curves revealed that, compared to PKM2(WT) [$k_{cat} = (538.4 \pm 18.0) s^{-1}$] the rate of product turnover was decreased by approximately 1.7-fold for PKM2(R120) [$k_{cat} = (308.4 \pm 5.7) s^{-1}$] and PKM2(K270A) [$k_{cat} = (270.8 \pm 5.3) s^{-1}$] (**Fig. 3.5 A**). Moreover, activity of PKM2(R72A) was almost entirely abolished, evidenced by a 63-fold decrease in the apparent rate of substrate turnover [$k_{cat} = (8.5 \pm 1.4)\ s^{-1}$] (**Fig. 3.5 A**).

To measure the inactivating effect of introducing catalytically inactive protomeric units into a hetero-oligomeric assembly with PKM2(WT), it was first necessary to establish that PKM2(R72A) did not disrupt oligomerisation. First, microscale thermophoresis (MST) measurements of fluorescein-labelled PKM2(WT) were performed with titrated amounts of unlabelled PKM2(WT), to measure PKM2(WT) oligomeric association. A concentration-dependent change in PKM2 thermophoresis was detected, from which an apparent oligomerisation constant (K_D^{oligo}) of (90.8 ± 5.7) nM was estimated for PKM2(WT) (Table 3.1). The apparent oligomerisation constant of PKM2(WT) was found to be unchanged upon addition of saturating concentrations of FBP, yielding a K_D^{oligo} of (97.9 ± 1.8) nM (Table 3.1).

MST measurements of labelled-PKM2(WT) were repeated with titrated amounts of PKM2(R72A), revealing a 4-fold increase in the K_D^{oligo} both in the absence [(410 ± 20) nM] and presence [(390 ± 40) nM] of saturating concentrations of FBP (**Fig. 3.5 B**) and **Table 3.1**). Despite the reduction in the apparent K_D^{oligo} for [WT]PKM2 and PKM2(R72A) association, a similar maximal thermophoretic behaviour of the labelled-PKM2(WT) at high concentrations of the mutant suggested that the catalytically-dead variant formed hetero-oligomeric assemblies with the wild-type, albeit with marginally weaker binding.

Titrant	Ligands	K_D^{oligo} (nM)
PKM2(WT)	None	90.8 ± 5.7
PKM2(WT)	2 μM FBP	97.9 ± 1.8
PKM2(WT)	2 μM Tepp-46	54.6 ± 2.1
PKM2(WT)	5 mM ADP	75.8 ± 6.4
PKM2(WT)	10 mM PEP	52.1 ± 10.1
PKM2(R72A)	None	410.0 ± 20.0
PKM2(R72A)	2 μM FBP	390.0 ± 40.0

Table 3.1: ***Dissociation constants for PKM2 self-association and hetero-association with PKM2(R72A), under various liganded conditions.*** *All measurements were made by titrating fluorscein-labelled PKM2(WT) with either unlablled PKM2(WT) or unlabelled PKM2(R72A).*

Titration of PKM2(WT) with PKM2(R72A) results in a non-linear decay in enzyme activity

Next, the mutant doping assay was performed whereby stoichiometric mixtures of PKM2(R72A) and PKM2(WT) were assayed for pyruvate kinase activity. PKM2(WT)-PKM2(R72A) mixtures were prepared at a total protein concentration of 5 μM, an order of magnitude in excess of the measured K_D^{oligo}, thereby circumventing the reduced affinity of the mutant for the wild-type. A linear decay in activity was observed over a range of substrate concentrations, with increasing $\frac{[PKM2(R72A)]}{[PKM2(WT)]}$, in the absence of added FBP (**Fig. 3.5 C**). Conversely, addition of saturating concentrations of FBP resulted in a non-linear decay in enzyme activity of the $\frac{[PKM2(R72A)]}{[PKM2(WT)]}$ mixtures (**Fig. 3.5 D**), suggesting an FBP-induced cooperative effect between the PKM2 subunits.

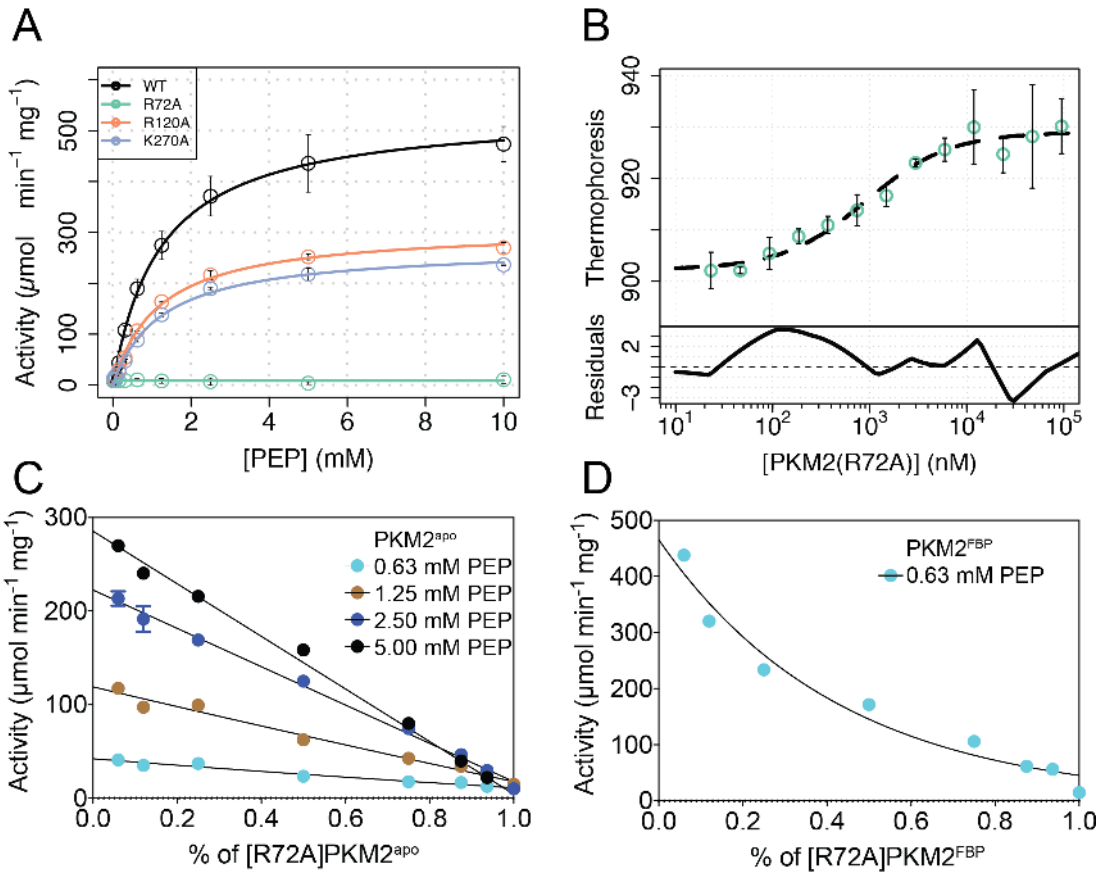

Figure 3.5: ***A mixture of wild-type and catalytically-dead PKM2 has a non-linear decay in activity.*** ***(A)*** *Initial velocity measurements were recorded over a range of PEP concentrations at a constant concentration of 5 mM ADP at 310 °C for PKM2(WT) (black) and active-site mutant variants PKM2(R72A) (teal), PKM2(R120A) (orange) and PKM2(K270A) (blue). The mean and standard deviation of four separate experiments are shown, for each protein variant. Kinetic parameters were estimated from Michaelis-Menten fits of the rate curves.* ***(B)*** *Microscale thermophoresis measurements were acquired for titrations of fluorescein-labelled PKM2(WT) with PKM2(R72A).* ***(C)*** *Mixtures of PKM2(WT) with PKM2(R72A) were assayed for enzyme activity over a range of substrate concentrations in the absence and* ***(D)*** *in the presence of 2 µM FBP.*

3.3.2 A numerical framework for modelling oligomeric enzyme cooperativity

The non-linear decay in enzyme activity of mutant-doped PKM2, upon FBP addition (Section 3.3.1), suggested a ligand-dependent inter-subunit cooperativity. To rationalise this observation, a numerical framework was constructed whereby the inter-subunit cooperativity of a tetrameric protein could be simulated, and compared to experimental data.

To this end, a model was generated to consider the enzyme activity resulting from the mixture of active PKM2(WT) oligomers (A) and inactive PKM2(R72A) (I) oligomers. We assumed that the measured association between PKM2(WT) and PKM2(R72A) formed tetramers and that the composition of the tetramers was purely stochastic. Moreover, it was posited that the mixture of A and I was defined by the ratio of the concentrations of A and I.

Given the above assumptions, five configurations of hetero-tetramers resulting from the mixture of PKM2(WT) and PKM2(R72A) are possible, not counting permutations under the assumption of a stochastic association: AAAA, AAAI, AAII, AIII, IIII. The concentration of A is given by a and the concentration of I is given by i. Therefore the ratio between A and I is:

$$r = \frac{a}{i} \tag{3.6}$$

The total concentration is trivially given by:

$$C = a + i \tag{3.7}$$

A final assumption made in this model was that the enzyme activity of A is 1 AU (*activity unit*) and that the corresponding activity of I is equal to 0 AU. Additionally, under conditions where A and I catalytic sites function in complete isolation (*ie.* no inter-subunit cooperativity), the

enzyme activity (ea) of the mixtures are given by the following expression:

$$ea(AAAA) = \frac{4}{3} \cdot ea(AAAI) = 2 \cdot ea(AAII) = 4 \cdot ea(AIII) \tag{3.8}$$

$$ea = \sum_{n=1}^{4} n(A + I) \tag{3.9}$$

From this, the probability of observing a fully active tetrameric species from a sequential mixture of a and i is given by a binomial distribution:

$$P(A) = \left[\binom{n}{A} a^A (1-a)^{n-A} \right]^{AC} \tag{3.10}$$

where AC is the allosteric coefficient, an added exponential term, defined by the strength of inter-protomeric cooperativity between individual subunits in the tetramer assembly (**Fig. 3.6 A**).

The mixture model in Equ. 3.10 was simulated over a range of increasing values of AC. Where $AC = 1$, a linear decrease in activity was observed (**Fig. 3.6 B**). Conversely, where $AC > 1$ a non-linear decay in activity of the model was observed (**Fig. 3.6 C**). The magnitude of the non-linear decrease in activity was found to be enhanced, with greater pre-defined values for AC (**Fig. 3.6 C**).

3.3.3 FBP binding induced inter-subunit cooperativity under conditions of limiting substrate concentrations

The mixture model of enzyme activity was applied towards analysing initial velocity measurements resulting from ratio-metric mixtures of PKM2(WT) and PKM2(R72) by performing a non-linear regression of the activity data to Equ. 3.10. Solutions for the allosteric coefficient (AC) were computed using a non-linear least squares procedure, as a means of quantifying

the strength of inter-protomeric cooperativity within hetero-oligomeric assemblies of PKM2.

The linear decay in enzyme activity, in the absence of any added FBP, supported the hypothesis that protomers within PKM2 oligomers were acting largely in isolation, with no detectable cooperativity between the individual catalytic pockets. This expectation was reflected in the solution to the AC, which was found to be $\simeq 1$ (**Fig. 3.6 D**). In contrast, the observed non-linear decay in enzyme activity, upon addition of saturating concentrations of FBP resulted in an increase in the solution for the allosteric coefficient. Moreover, this non-linear effect, and as such the allosteric coefficient strength, was found to increase at concentrations of substrate below the K_M^{PEP} (**Fig. 3.6 D**). The correlation of the AC with the concentration of phosphoenolpyruvate suggested a mechanism whereby FBP binding acts to increase the cooperativity between individual subunits under conditions where the substrate is limiting. Conversely, when the substrate concentrations are near-saturating, FBP-induced inter-subunit cooperativity is diminished, regulate the activity of PKM2.

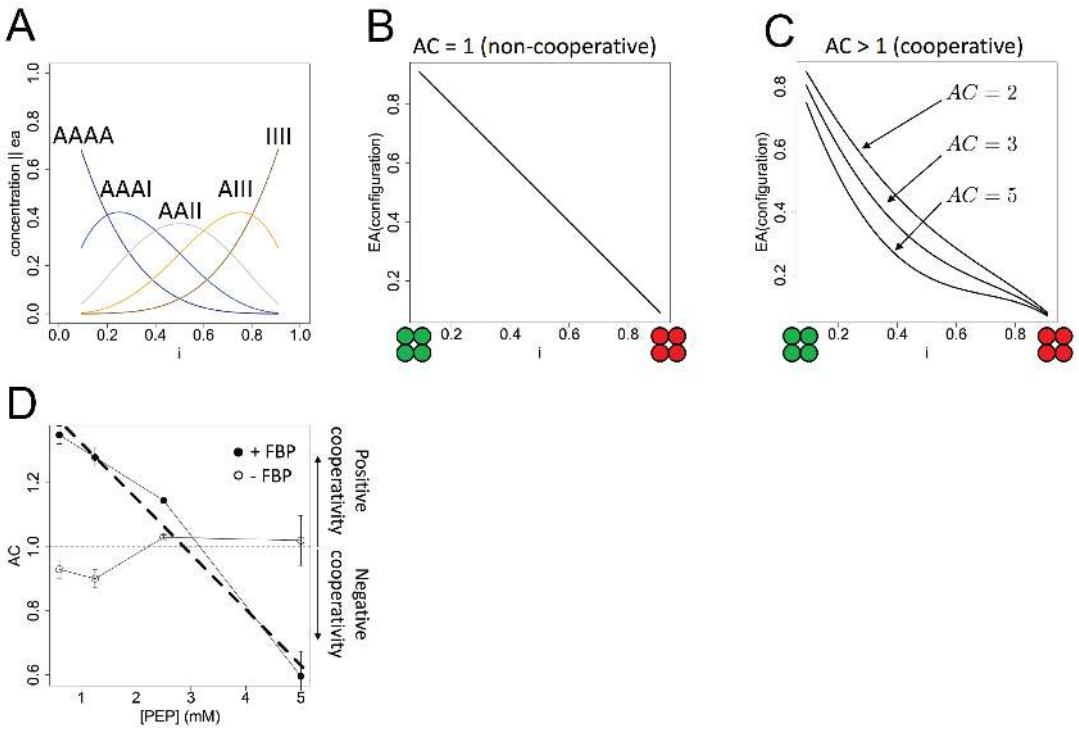

Figure 3.6: ***FBP binding results in inter-subunit cooperativity as defined in a numerical model.*** ***(A)*** *A probability distribution for the five model hetero-oligomers AAAA, AAAI, AAII, AIII and IIII over a range of mixtures of active and inactive protein (i).* ***(B)*** *Simulation of the decay in enzyme activity of a non-cooperative system.* ***(C)*** *Simulation in the decay in enzyme activity for a cooperative system, where increasing allosteric coefficients (AC) model the magnitude of inter-subunit cooperativity.* ***(D)*** *Solution for the allosteric cooperativity coefficient for WT:R72A mixtures in the absence of any added FBP (empty) and in the presence of 2 μM FBP (solid).*

3.4 FBP and amino acids combinatorially regulate PKM2 activity

In addition to the functional effects of FBP on PKM2, demonstrated here, several amino acids compete for a common allosteric pocket[19,20,22] to regulate PKM2 enzyme activity. The regulatory effects of amino-acid binding *per se* is disputed; some studies have reported allosteric effects on substrate affinity[18,19,59,71], while others report effects on the rate of substrate turnover[20,22]. Moreover, little is known about how simultaneous binding of FBP and amino acids concurrently regulate PKM2.

3.4.1 Phe and Ser *per se* are K-type modulators of PKM2 catalytic activity

To this end, we started by measuring the functional effects of an activating amino acid (L-serine; Ser) and an inhibitory amino acid (L-phenylalanine; Phe) on PKM2 enzyme catalysis *per se*. Phe and Ser were chosen because their binding had been structurally resolved[19,22], though at the time of writing a subsequent study by Yuan *et al.* (2018)[20] reported further structures of PKM2 in complex with Ala, Trp and His. PKM2 enzyme activity was measured in the absence and in the presence of Ser and Phe. Consistent with previously published findings, we found Phe addition to increase the Michaelis-Menten constant for the substrate K_M^{PEP} to (7.09 ± 1.58) mM, without a detectable effect on the product turnover number [k_{cat} = (324.7 ± 23.0) s^{-1}] (**Fig. 3.7**). Conversely, addition of Ser resulted in a decrease in the K_M^{PEP} to (0.22 ± 0.03) mM, without a significant change in the rate of product turnover (**Fig. 3.7**). These results show that, in agreement with several existing studies[18,19,59,71], Ser and Phe regulate PKM2 in a K-type manner *per se* by modulating the apparent substrate affinity for PEP, without changing the apparent rate of substrate turnover.

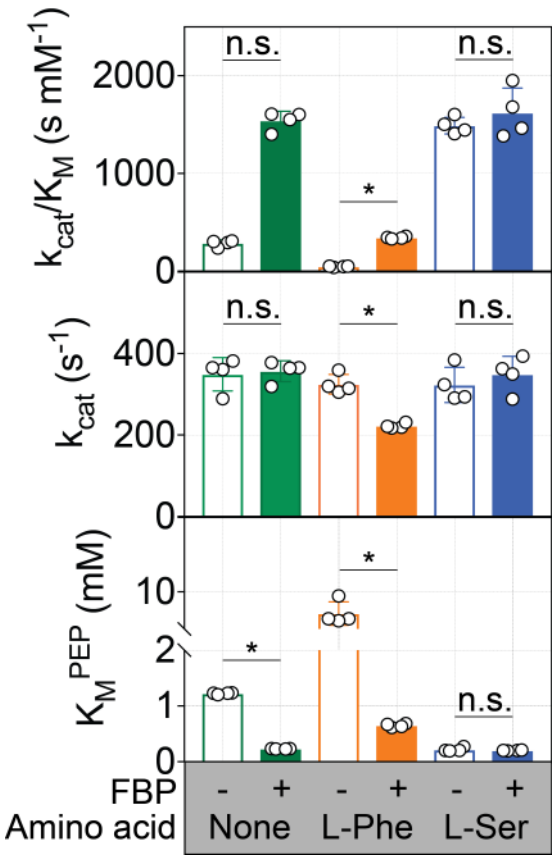

Figure 3.7: ***FBP and amino acids have distinct combined roles in regulating PKM2 activity.*** *Steady-state kinetic parameters of recombinant PKM2 were measured with a spectrophotometric assay coupled to lactate dehydrogenase over a range of concentrations of the substrate phosphoenolpyruvate at 37 °C. Kinetic parameters K_M and k_{cat} were estimated in the absence of added ligands (empty green bars) and in the presence of 2 μM FBP (solid green), 400 μM Phe (empty orange), 400 μM Phe and 2 μM FBP (solid orange), 200 mM Ser (empty blue) and 200 mM Ser and 2 μM FBP (solid orange). Measurements were independently repeated four times. Significance was assessed using a Wilcoxon rank-sum test. Asterisk (*) marks significant changes (p-value < 0.05). Data are shown in a tabular form in* ***Table 3.2****.*

3.4.2 The mechanism of phenylalanine inhibition of PKM2 catalysis is distinct depending on FBP binding.

In order to investigate the inter-dependence between amino acid and FBP regulation, PKM2 enzyme activity was measured following addition of FBP and either Ser or Phe. Addition of Ser, subsequent to FBP did not produce detectable changes to the K_M^{PEP} and the k_{cat} (**Fig. 3.7**), showing that FBP and Ser do not synergistically activate PKM2.

In the presence of FBP, Phe reduces the substrate affinity and the maximal velocity of PKM2 catalysis

Next, PKM2 activity was measured in the presence of FBP and Phe. We found that in the presence of FBP, Phe significantly decreased the k_{cat} of PKM2 in a dose-dependent manner by 57 % from 349.3 s^{-1} at [Phe] = 0 μM, to 222.8 s^{-1} at [Phe] = 400 μM. In contrast, no Phe-dependent change to the k_{cat} was detected in the absence of FBP (**Fig. 3.7**). In addition to the observed reduction in the product turnover number, Phe addition reduced the operational substrate binding affinity from 0.22 mM, in the absence of Phe and saturating FBP, to (0.65 $\pm$ 0.03) mM at [Phe] = 400 μM (**Fig. 3.7**. Taken together, these data revealed that Phe addition, in the presence of FBP, produces a distinct inhibitory effect on PKM2 activity compared to the effects of Phe addition *per se*.

Phe *per se* acts as a hyperbolic specific inhibitor of PKM2

In order to assign the mechanism of Phe inhibition *per se*, the dependence of the enzyme kinetic constants K_M, k_{cat} and $(\frac{k_{cat}}{K_M})^{app}$ on the concentration of phenyalanine were analysed (as described in Methods Section 2.3.4) (**Fig. 3.8 A-C**). Phenylalanine was found to act as a hyperbolic specific inhibitor of PKM2 in the absence of FBP (**Fig. 3.8 D**). Formation of the enzyme-effector (EX) complex was favoured over the enzyme-substrate complex (ES) due to an affinity of the enzyme for Phe ($K_X = 0.019\,mM$), which was approximately 14-times greater than the measured effective substrate affinity ($K_S = 0.250\,mM$). However, the affinity of the substrate for the EX complex was found to be very weak ($\alpha K_S > 20\,mM$),

thus sequestering the enzyme in the EX complex, leaving less enzyme available for the ES and ESX complexes, to facilitate turnover of the product. Notably, increasing concentrations of phenylalanine had little effect on the coefficient β and so the rate of product turnover (k_2) between the ES and ESX complexes remained unchanged.

In the presence of FBP, Phe *per se* acts as a hyperbolic mixed inhibitor of PKM2

The dependence of the enzyme kinetic constants K_M, k_{cat} and $(\frac{k_{cat}}{K_M})^{app}$ on the concentration of phenyalanine was next analysed in the presence of saturating amounts of FBP (**Fig. 3.8 E-G**). In the presence of saturating FBP, the effective substrate binding constant to ^{FBP}E ($K_S = 0.170$ mM) was found to be approximately three-times greater than the effector binding constant ($K_X = 0.540$ mM), marginally favouring the formation of the ^{FBP}ES complex over the ^{FBP}ESX complex (**Fig. 3.8 H**). Nevertheless, Phe was found to have a significant effect on the effective substrate affinity of the ^{FBP}EX complex ($\alpha K_S = 1.130$ mM), compared to substrate binding to ^{FBP}E, thereby acting to sequester a fraction of FBP-bound enzyme away from substrate binding. Moreover, Phe addition was found to significantly reduce the β coefficient resulting in a reduction of the rate of product turnover from the ^{FBP}ESX complex ($\beta k_2 = 174.4\ s^{-1}$), compared to the rate of product turnover from the ^{FBP}ES complex ($\beta k_2 = 459.0\ s^{-1}$). Therefore, the inhibitory effect of Phe binding to $PKM2^{FBP}$ could be accounted for by a mixed mechanism of both reduced substrate binding *and* reduced product turnover, and as such could be assigned as a hyperbolic mixed (predominantly specific) allosteric inhibitor of $PKM2^{FBP}$ (**Fig. 3.8 H**).

Together, these data revealed that FBP mechanistically alters the mode of PKM2 inhibition by phenylalanine from a hyperbolic specific inhibitor, to a mixed hyperbolic inhibitor. When FBP is absent, Phe acts exclusively as a K-type inhibitor whereas, in the presence of saturating concentrations of FBP, Phe addition acts on both the K_M^{PEP} and the k_{cat}, as a mixed V- and K-type inhibitor.

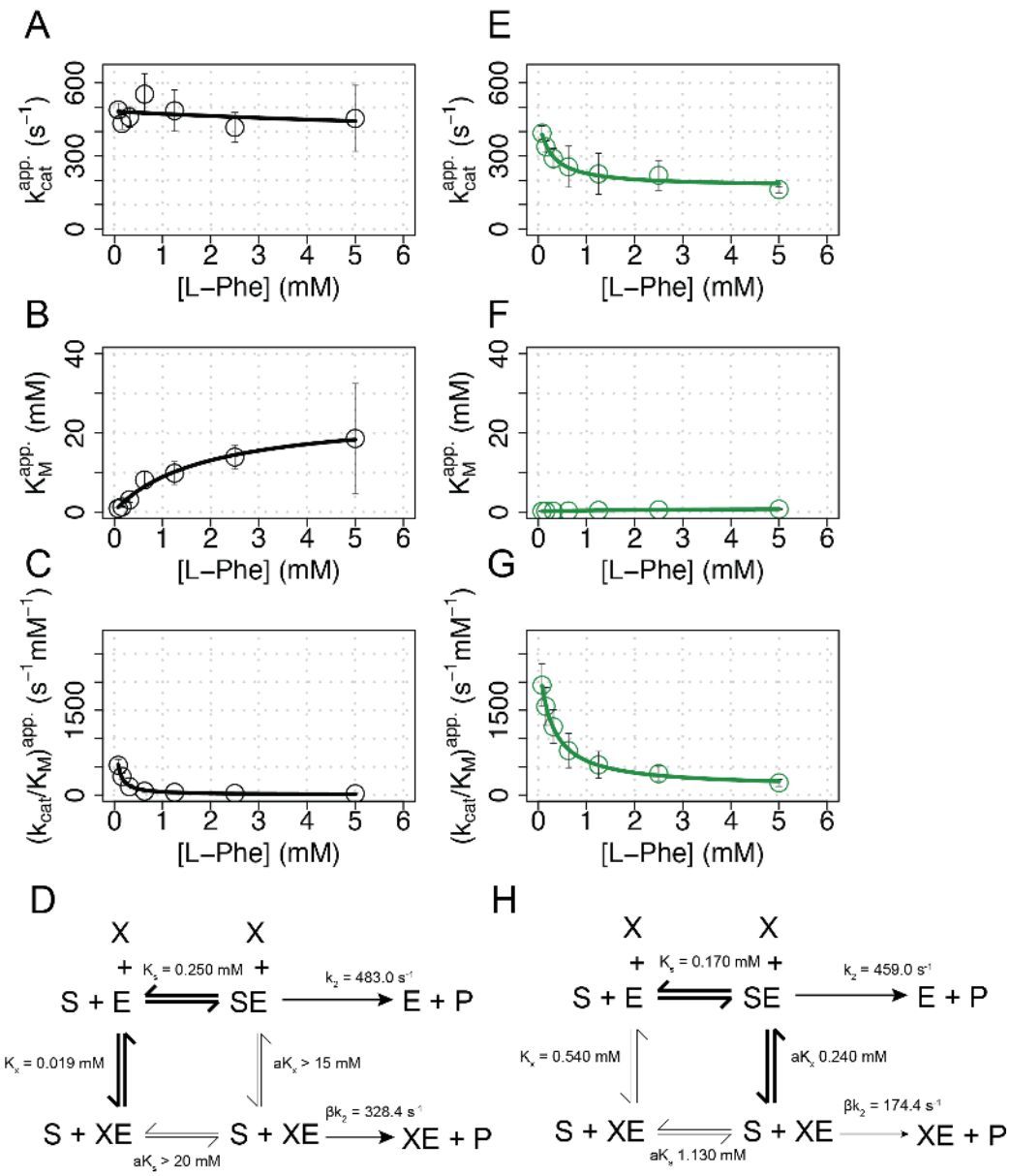

Figure 3.8: ***Phenylalanine inhibition follows one of two distinct mechanisms, depending on the presence of FBP.*** *Recombinant PKM2 activity was measured at varying concentrations of PEP and a constant concentrations of 5 mM ADP and 5 nM PKM2, at 37 °C. Initial velocities were performed in the presence of a range of Phe concentrations between 0.08 mM and 5 mM. Rate curves were fitted to Michaelis-Menten kinetic models, from which the parameters k_{cat}, K_M^{PEP} and $\frac{k_{cat}}{K_M^{PEP}}$ were computed for Phe inhibition of $PKM2^{apo*}$ (black;* ***A****,* ***B*** *and* ***C****). **(D)** The dependence of the three steady-state kinetic parameters k_{cat}, K_M^{PEP} and $\frac{k_{cat}}{K_M^{PEP}}$ were fit to the steady-state modifier-rate equations defined in the text.* ***(E-H)*** *An analysis of phenylalanine inhibition was repeated as in (A-D), in the presence of 2 μM FBP (saturating amounts).*

Ligand	K_M^{PEP} (mM)	k_{cat} (s^{-1})	$\frac{k_{cat}}{K_M^{PEP}}$ $(s^{-1} \cdot mM^{-1})$
Apo$*$	1.22 ± 0.02	349.3 ± 40.9	285.6 ± 34.1
Phe	7.08 ± 1.58	324.7 ± 23.9	46.8 ± 6.0
Ser	0.22 ± 0.04	323.1 ± 43.2	1489.8 ± 84.7
FBP	0.23 ± 0.04	356.7 ± 25.7	1540.4 ± 96.9
FBP and Phe	0.65 ± 0.03	222.3 ± 6.3	342.1 ± 11.1
FBP and Ser	0.20 ± 0.04	348.7 ± 44.3	1620.0 ± 253.6

Table 3.2: ***Steady-state Michaelis-Menten kinetic parameters for*** $PKM2^{apo*}$***, and following the addition of FBP, Phe and Ser.***

3.4.3 Ligand-induced changes in affinity cannot explain the changed phenylalanine-inhibition of PKM2 by FBP

Results in Sections 3.4.1 and 3.4.2 revealed that FBP addition changes the inhibition kinetics of Phe. It was unclear whether this was a result of a combinatorial effect of both ligands on the bind affinity of the other, or whether the result reflected a *bona fide* functional cross-talk between the two ligands. Indeed, the effective Phe binding constant to $PKM2^{apo*}$ [K_X^{apo*} = (0.019 ± 0.077) mM] was found to be significantly different compared to that of Phe binding to $PKM2^{FBP}$ [K_X^{FBP} = (0.540 ± 0.06) mM] (**Fig. 3.8 D** and **H**). Nevertheless, since a simplified single-substrate-single-enzyme mechanism was assumed, it was unclear whether the changed (K_X) reflected a changed binding affinity or whether a more complex interaction between PKM2, phosphoenolpyruvate and adenosine diphosphate (the second substrate not considered in the above mechanism) was at play. To this end, the mutual effects of FBP and amino acids on binding to PKM2 was investigated.

We first estimated the apparent affinities of each ligand in the presence of saturating amounts of the other. In the presence of saturating concentrations of either Phe or Ser, the K_D^{FBP} decreased, however this was found not to be statistically significant (p = 0.150 and p = 0.054, respectively) (**Fig. 3.9 A**).

Next, the binding of Phe and Ser to PKM2 was measured, in the absence and in the presence of FBP, using MST. Phe and Ser addition to PKM2 produced dose-dependent changes to the thermophoresis of labelled PKM2. The estimated binding affinites of Phe [(191.0 ± 86.2) μM) and Ser [(507.5 ± 218.8) μM) (**Fig. 3.9 B**), were unaffected by the addition of saturating amounts of FBP (**Fig. 3.9 C**).

Taken together, these measurements suggested that alteration in the binding affinity of one ligand by the presence of the other cannot account for the distinct FBP-dependent modes of PKM2 inhibition by Phe. Rather, the altered mechanism of Phe inhibition upon FBP binding

may reflect a mechanistic dependence between the two allosteric regulators.

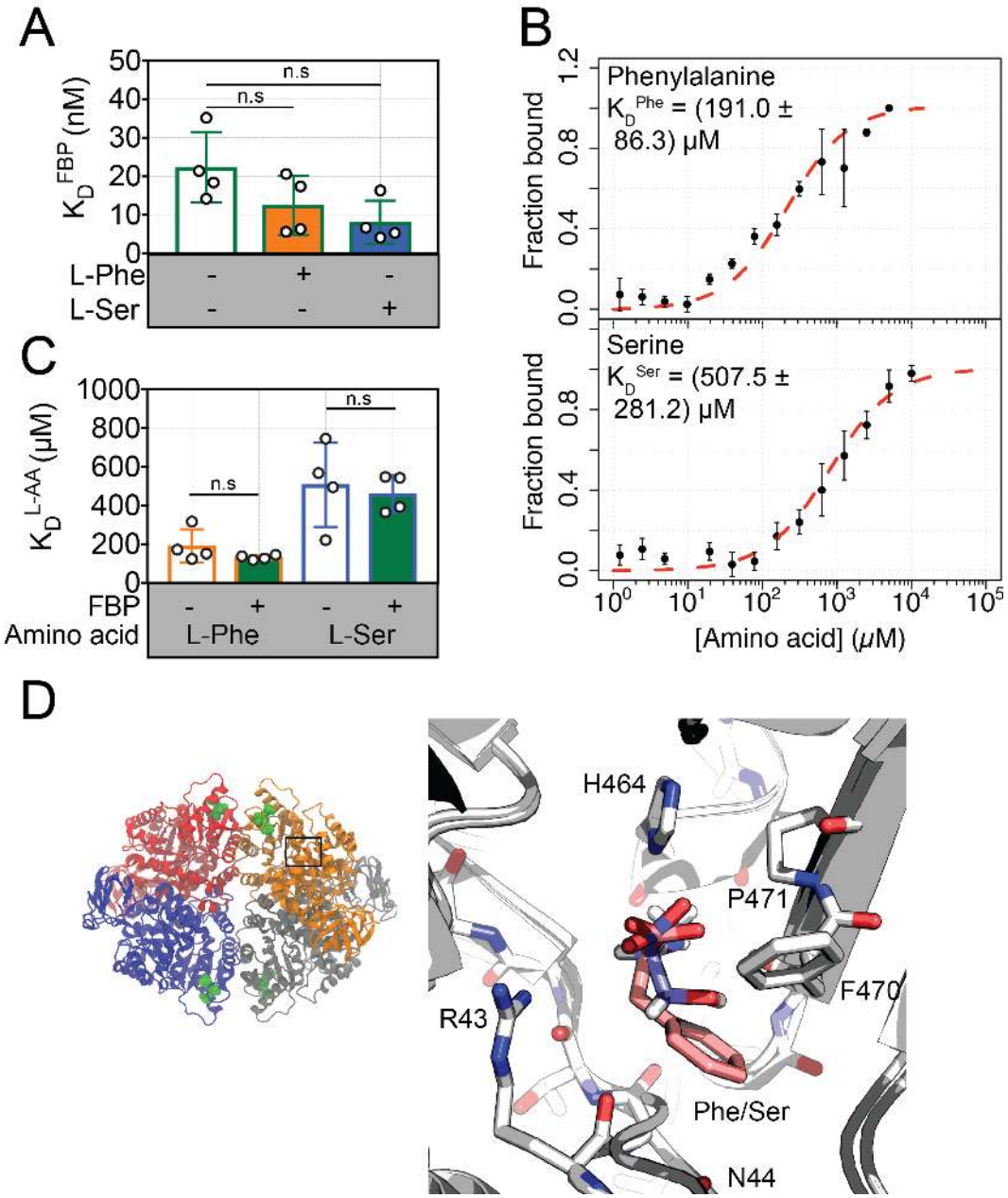

Figure 3.9: ***Amino acids and FBP do not reciprocally affect binding.*** ***(A)*** *FBP binding to PKM2 was measured using fluorescence emission spectroscopy (as described previously) in the absence of amino acids (green) and in the presence of 400 µM Phe (orange) and 200 mM Ser.* ***(B)*** *MST measurements of fluorescein-labelled PKM2 were performed to monitor serine (Ser) and phenylalanine (Phe) binding to PKM2. Labelled PKM2, at a constant concentration of 30 nM, was titrated with either Phe or Ser, and the thermophoretic properties of the resulting protein-ligand interaction was measured. The mean and standard deviations of four separate titrations are shown.* ***(C)*** *MST experiments were repeated for Phe (orange) and Ser (blue) in the presence of 2 µM FBP.* ***(D)*** *The overlapping modes of Phe and Ser binding to PKM2 are superimposed from two crystal structures*[19,22].

3.5 Simultaneous regulation of PKM2 by multiple ligands is relevant for enzyme regulation in cells

Results presented in this chapter demonstrate that FBP binds tightly to PKM2 with a high affinity, and that Phe attenuates FBP-induced activation by stabilising a distinct kinetic state with a reduced maximal velocity. The fraction of PKM2 bound to allosteric ligands in cells, and hence the ability of these ligands to exert their respective regulatory effects, is determined by their intracellular concentrations and their binding affinities. To investigate whether FBP and amino acids are likely to concurrently bind to PKM2 and whether this is relevant for PKM2 regulation, we sought to quantify the fractional saturation of PKM2 bound to FBP, Ser and Phe, in proliferating cancer cells. The cellular fraction of PKM2 bound to allosteric ligands is given by:

$$[PL] = \frac{(K_D + [P_{ic}] + [X_{ic}]) - \sqrt{(K_D + [P_{ic}] + [X_{ic}])^2 - 4([P_{ic}][X_{ic}])}}{2} \tag{3.11}$$

$$s = \frac{[PL]}{[P_{ic}]} \tag{3.12}$$

where $[P_{ic}]$ and $[X_{ic}]$ are the intracellular protein and metabolite concentrations, respectively.

3.5.1 Intracellular FBP binding to PKM2 is constitutive whereas Phe and Ser bind reversibly

The intracellular concentration of FBP is in excess of its binding affinity to PKM2

To this end, the range of intracellular concentrations of FBP, Ser and Phe in three human cancer cells lines of different tissue origin (HCT116, LN229 and SN12C), cultured in the presence of physiological (human blood serum) concentrations of glucose ($Gluc^+$). $[FBP]_{ic}$ varied between between 240-360 μM across the three cell lines (**Fig. 3.10 A**). This concentration range of FBP, under physiological glucose culture conditions, was in excess of the K_D^{FBP} be-

tween 11919- and 17755-fold. The fraction of cellular PKM2 bound to FBP, calculated using **Equ. 3.12**, found a fractional saturation of 0.99, which remained unchanged even in cells cultured in the absence of glucose ($Gluc^{-}$) (**Fig. 3.10 B**), despite a marked reduction in the $[FBP]_{ic}$ to between 20-70 μM across the three cell lines.

The intracellular concentrations of Ser and Phe constitute a dynamic range around their respective affinities for PKM2

Next, cells were supplemented with either 100 μM Ser and 500 μM Phe ($S^{100}F^{500}$) or 500 μM Ser and 100 μM Phe $S^{100}F^{500}$). Under physiological media concentrations, the intracellular concentrations varied between 200-320 μM for Phe and 2000-4000 μM for Ser, across the three cell lines (**Fig. 3.10 C**). The measured $[Phe]_{ic}$ and $[Ser]_{ic}$ were close to the apparent binding affinities for PKM2 [$K_D^{Phe} = (191.0 \pm 86.3)$ μM and $K_D^{Ser} = (507.5 \pm 218.2)$ μM]. The associated fractional saturation of PKM2 with Phe was between 0.53-0.75, and between 0.81 and 0.92 with Ser. The range of predicted fractional saturations in the complete absence of amino acids (aa^{-}) or 5x physiological concentrations was 0.05 and 0.18 for Phe and Ser, respectively, in culture conditions devoid of amino acids (aa^{-}). The fractional saturation of both Phe and Ser was found to increase to 0.90 and 0.98, respectively, in conditions of $S^{100}F^{500}$ and $S^{500}F^{100}$ (**Fig. 3.10 D**).

Estimations for the fraction of PKM2 bound to Ser and Phe did not take into account the fact that these two ligands compete for binding. Moreover, the calculation assumed that binding was unaffected by other amino acids, which have been shown to bind competitively at the same allosteric pocket[20]. As such, the predicted fractional saturation of Phe and Ser is likely to be an under-estimation of the true cellular value. Conversely, FBP binding was predicted to be constitutively saturated, evidenced by a predicted binding saturation approaching 1 in all conditions (**Fig. 3.10 D**).

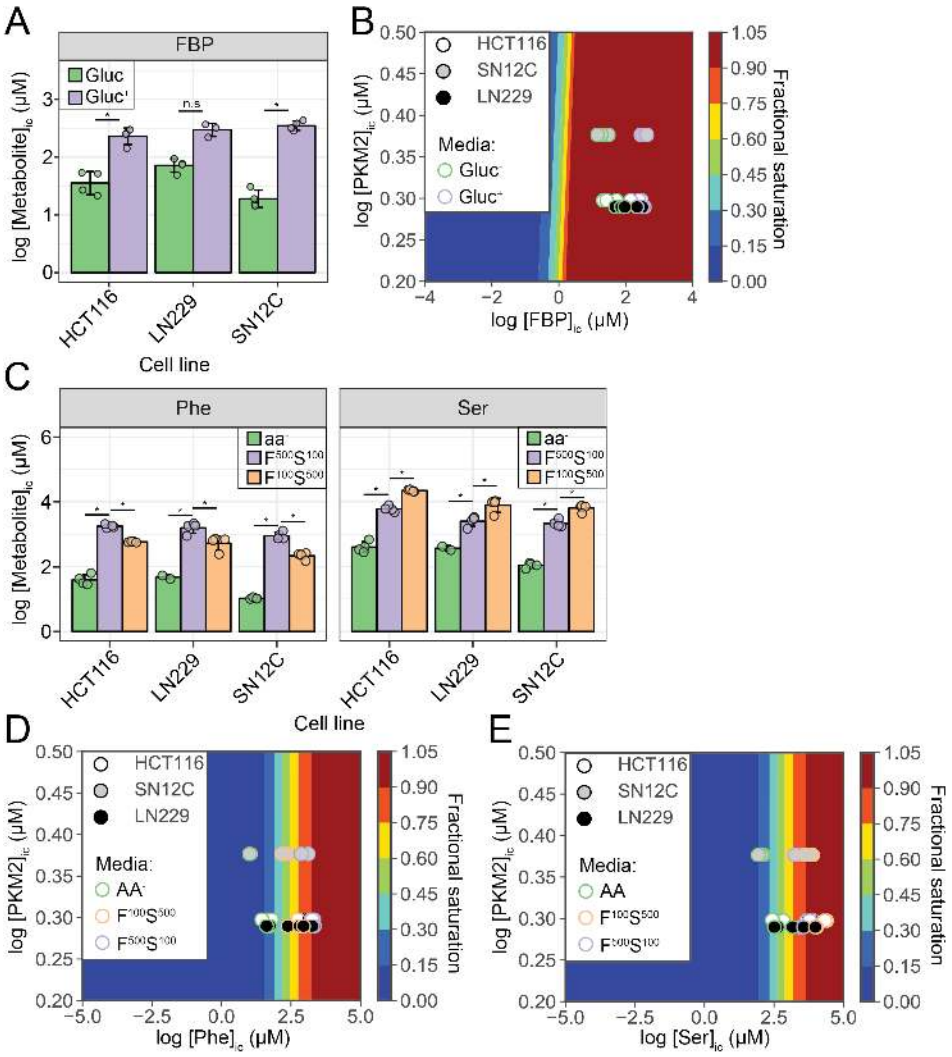

Figure 3.10: ***Cellular concentrations of PKM2 allosteric effectors reveal a physiological role for multi-ligand regulation.*** ***(A)*** *Intracellular concentrations of FBP were measured using liquid-chromatography mass spectrometry (LC-MS) from HCT116, LN229 and SN12C cells cultured in fully-fed conditions (RPMI containing 11 mM glucose; green bars) and glucose-deprived conditions (RPMI containing 0 mM glucose; purple bars) for 1 hour.* ***(B)*** *A phase diagram for intracellular FBP binding to PKM2 was computed over a range of [FBP] and [PKM2] values. Saturation of ligand-protein binding is represented by a colour scale; a fractional saturation approaching 0 indicating conditions under which PKM2 is fully apo∗ with respect to its interaction with FBP, and a fractional saturation equal to 1 indicates that each FBP binding site on the cellular pool of PKM2 is occupied. The predicted fractional saturation for each of the three cell lines (four technical replicates) are shown in the phase diagram. The intracellular concentration of PKM2 was measured for each cell line using a targeted proteomics approach (see Methods Section 2.5.2).* ***(C)*** *Intraceullar concentrations of Phe and Ser were measured in media devoid of amino acids (aa^-), containing 500 μM Phe and 100 μM Ser ($F^{500}S^{100}$) or 100 μM Phe and 500 μM Ser ($F^{500}S^{100}$). Phase diagrams were computed for phenylalanine* ***(D)*** *and serine* ***(E)****, as for FBP above.*

3.5.2 The addition of exogenous FBP does not change PKM2 activity in HCT116 cell lysates

Under conditions where FBP binding is constitutive, the intracellular activity of PKM2 would be predicted to be unresponsive to the addition of further amounts of exogenous FBP. To test this hypothesis, we assayed HCT116 cell lysates for PKM2 activity, following culture in glucose-fed and glucose-depleted media. Cells were gently lysed in hypotonic lysis buffer, in an attempt to maintain the intracellular FBP-PKM2 stoichiometry. It was observed that, while addition of 100 μM FBP to recombinant PKM2 had the effect of decreasing the K_M^{PEP} by 7-fold, very little change to PKM2 activity was apparent in HCT116 lysates upon addition of up to 1 mM exogenous FBP (**Fig. 3.11 A**).

Phe addition reduces the maximal velocity of PKM2 activity in cell lysates

Addition of physiological concentrations of exogenous Phe to HCT116 cell lysates resulted in an apparent inhibition of PKM2 activity, and a dose-dependent reduction in the $\frac{k_{cat}}{K_M^{PEP}}$ (**Fig. 3.11 B**). It was observed that a Phe dose-depedent reduction in the maximal velocity was the main kinetic determinant in the inhibition of the $\frac{k_{cat}}{K_M^{PEP}}$ of PKM2. No detectable change to the K_M^{PEP}, however, was apparent upon addition of Phe to HCT116 lysates, similar to the observed mechanism of Phe inhibition of recombinant $PKM2^{FBP}$. Furthermore, the addition of exogenous molar concentrations of Ser out-competed the inhibitory effect of Phe, restoring the maximal velocity to 'FBP-bound' levels (**Fig. 3.11 C**).

Binding of PKM2 to PEP in cells is predicted to be sub-saturating

Intracellular concentrations of the catalytic substrate phosphoenolpyruvate (PEP) were measured in the low micro-molar range between (9.6 $\pm$ 3.5) μM and (2.3 $\pm$ 1.3) μM under glucose-fed and glucose-deprived conditions, respectively (**Fig. 3.12 A**). The estimated K_A^{PEP} was therefore between approximately 24-fold and 100-fold in excess of $[PEP]_{ic}$, suggesting that a very low fraction of the protein was bound to its substrate in cells ($s_{PEP} < 0.04$) (**Fig. 3.12**

B). Conditions where the substrate is limiting give further credence to the relevance of Phe and Ser competitively modulating the rate of product turnover, in the context of constitutive FBP binding. Nevertheless, possible channelling of the substrate and/or differential PEP sequestration in different cellular compartments, could not be discounted.

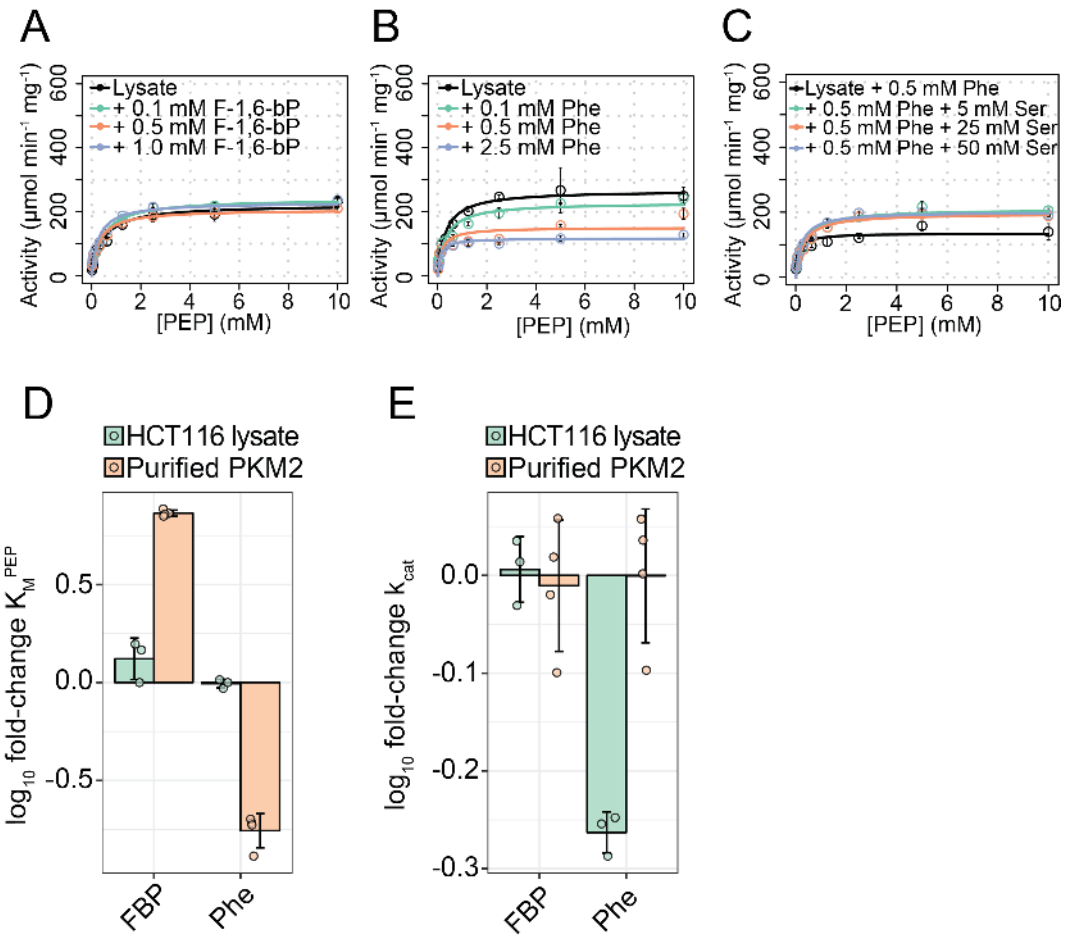

Figure 3.11: ***PKM2 activity from HCT116 lysates is not further activated by addition of exogenous FBP.*** ***(A)*** *PKM2 activity in HCT116 cells cultured under glucose-fed conditions was assayed, in the presence of increasing concentrations of exogenously added FBP. Initial velocities were measured over a range of PEP concentrations.* ***(B)*** *Activity measurements were repeated in presence of increasing concentrations of exogenous phenylalanine and* ***(C)*** *serine. The fold-change in the Michaelis-Menten constants* ***(D)*** K_M^{PEP} *and* k_{cat} *measured from PKM2 activity of recombinant PKM2 (orange) and PKM2 activity in HCT116 cell lysates (teal). The fold-change is plotted between the 'apo∗' condition and upon addition of either 2 µM FBP or 400 µM Phe.*

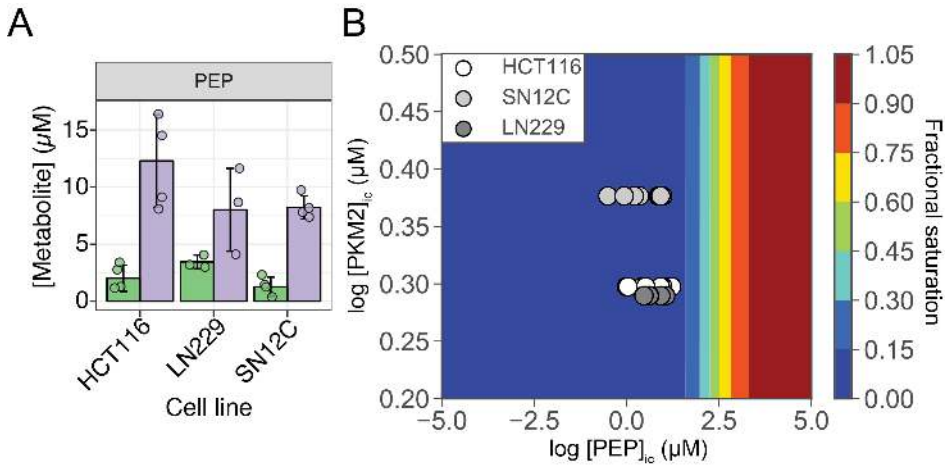

Figure 3.12: ***Intracellular concentrations of phosphoenolpyruvate are predicted to be constitutively limiting for PKM2 binding.*** ***(A)*** *Intracellular concentrations of phosphoenolpyruvate were determined in three cancer cell lines under 11 mM glucose-cultured (teal) and glucose-deprived (red) conditions.* ***(B)*** *A phase diagram of phosphoenolpyruvate binding to PKM2. A binding constant of 0.22 mM was used, as determined from PKM2 activity measurements under conditions of FBP binding.*

3.5.3 High milli-molar phospho-peptide concentrations are required to reversibly displace FBP binding

In this chapter, it is shown that competition between Phe and Ser regulating the rate of $PKM2^{FBP}$ product turnover. The observed V-type effect of Phe and Ser is dependent on saturating FBP concentrations. We therefore reasoned that the combined regulatory effects of amino acids and FBP could be perturbed upon phosphopeptide-induced displacement of FBP binding. Nevertheless, exploring the stoichiometry of peptide binding to PKM2 in cells is complicated by the likelihood that multiple peptide binding motifs bind with varying affinities. Therefore, an accurate quantification of the fraction of PKM2 bound to phospho-peptides would require a comprehensive phospho-proteomics screen combined with a high-throughput peptide affinity effort, which was beyond the scope of this thesis.

Instead, to estimate the intracellular concentration of PKM2-binding phosphopeptide, required to reduce the binding affinity of FBP to levels which facilitate a non-constitutive phenotype, Equ. 3.12 was computed for a range of simulated phosphopeptide concentrations. It was assumed that all phosphopeptide molecules have the same binding affinity for PKM2, and that they all displace FBP with the same kinetics as measured in recombinant protein (see Section 3.2.4). We found that phosphopeptide concentrations of approximately 40 mM were required to reduce the fractional saturation of FBP to 0.5 (**Fig. 3.13**).

Together, these observations supported the hypothesis that, during steady-state cell proliferation under cell culture conditions, FBP is constitutively bound to PKM2. Under these conditions, amino acids can reversibly bind and thereby regulate the enzyme activity of PKM2 by modulating its maximal velocity, in the background of post-translational modifications that are likely to occur to PKM2 in cells.

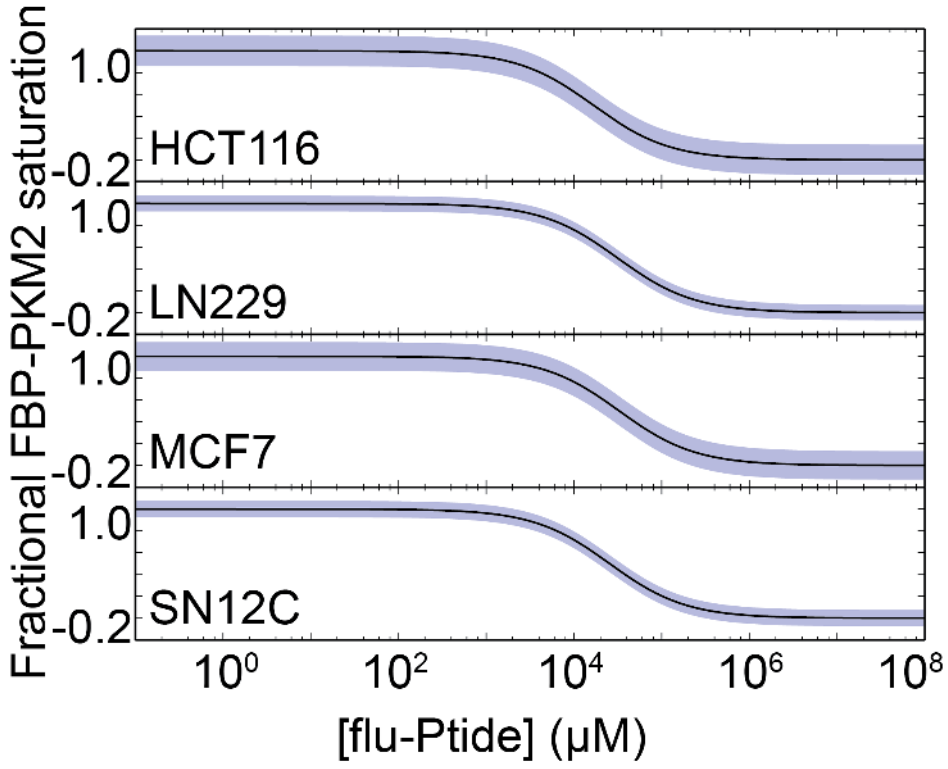

Figure 3.13: ***Milli-molar concentrations of phosphorylated peptide is required to out-compete FBP binding in cells.*** *The fraction of PKM2 bound to FBP in HCT116, SN12C, MCF7 and LN229 cell lines was computed over a range of phospho-peptide concentrations.*

3.6 Conclusion

Evidence, presented in this chapter, derived from measurements of PKM2 activity suggests that allosteric effectors can modulate the velocity and substrate affinity through a complex series of intra- and inter-protomeric couplings. Moreover, simultaneous regulation of PKM2 by FBP and amino acids is likely to occur over a large range of cell culture conditions. A major mechanism that has been suggested to result from allosteric effector binding is the regulation of the oligomeric state of PKM2. We therefore wanted to understand if, and to which extent, the findings of this chapter could be explained by ligand-induced oligomerisation of PKM2. We next turned to native mass spectrometry to investigate the potential of allosteric regulators to effect changes in the oligomeric state of PKM2.

Chapter 4

Native mass spectrometry reveals the structural basis of distinct ligand modes of action.

4.1 Introduction

The oligomeric state adopted by PKM2, and how ligand binding affects changes to the monomer-dimer-tetramer equilibrium, remains controversial. PKM2 forms a homotetrameric complex in crystal structures, both in the absence and in the presence of various co-crystalised allosteric ligands and substrates[17,19,20,22,23,55]. Solution-phase chromatographic measurements of PKM2 size, however, have shown that PKM2 forms an equilibrium of monomers and tetramers[17,20,22,24,70], with some studies reporting the existence of dimers[21,62], in which the tetramer has high enzymatic activity whereas monomers and dimers display lower levels of activity[17,33,65,70,71]. Accordingly, the regulation of PKM2 oligomerisation by various allosteric ligands in solution[17,20–22,62,70,71] and in the gas phase[63], has been previously reported as a mechanism by which PKM2 activity can be modulated. In addition to oligomerisation, several studies of PKM2 and its homologues have implicated ligand-induced conformational changes as contributing to the inactive-active state transition[20,22,55,66,72–74,77]. What is unclear, how-

ever, is the relative contribution of ligand-induced oligomerisation and conformational changes towards regulation of PKM2 enzyme activity.

Measurements of the steady-state enzyme kinetics of PKM2, in Chapter 3, found that allosteric regulators FBP, Ser and Phe act *per se* in a K-type mechanism to modulate the effective substrate affinity of PKM2 for PEP. Moreover, it was found that FBP binding alters the mode of phenylalanine inhibition from a hyperbolic-specific to a hyperbolic-mixed mechanism, suggesting a mechanistic cross-talk between the two ligands. Given that modulation of the oligomeric state of PKM2 has been previously demonstrated as an important regulatory mechanism, we sought to investigate whether ligand-induced activity changes could be explained by PKM2 oligomerisation and/or conformational transitions.

In this chapter, a characterisation of PKM2 oligomerisation in the gas phase is presented. Two mass spectrometry techniques, nano-electrospray ionisation mass spectrometry (nESI-MS) and ion-mobility coupled to nESI-MS (IM-MS), were used to explore the distribution of PKM2 oligomers and ensemble of conformations, respectively in the native-state.

4.2 FBP binding regulates oligomerisation of PKM2

4.2.1 PKM2 forms a mixture of oligomeric states

To mass-resolve the oligomeric states of PKM2 in the gas phase, native spectra of PKM2 in the absence of any added ligands were obtained. As was detailed in Chapter 3, preparations of PKM2 co-purified with sub-stoichiometric amounts of FBP. All experiments were performed with a single preparation of PKM2 containing less than 25 % of its sites pre-bound to FBP. Experiments performed on purified PKM2 in the absence of any added ligands, though containing quantified sub-stoichiometric amounts of FBP, are denoted with an asterisk (Apo$*$).

The native spectrum of PKM2 showed a mixture of monomers, dimers and tetramers at an approximate ratio of 1:7:10 (**Fig. 4.1 A**). To investigate the concentration-dependence of PKM2 oligomerisation, additional native spectra of PKM2 were acquired at protein concentrations between 1 μM and 70 μM. A concentration-dependent effect on the relative intensities of the three oligomeric species was observed, favouring a higher proportion of tetramers in the more concentration protein samples at the expense of the dimeric species (**Fig. 4.1 A** and **B**). Additionally, the relative intensities of the monomeric charged-state species were found to increase towards higher protein concentrations. This may suggest an inherent instability of the dimeric species at higher protein concentrations, which pushes the equilibrium of oligomeric states towards the tetrameric and monomeric states.

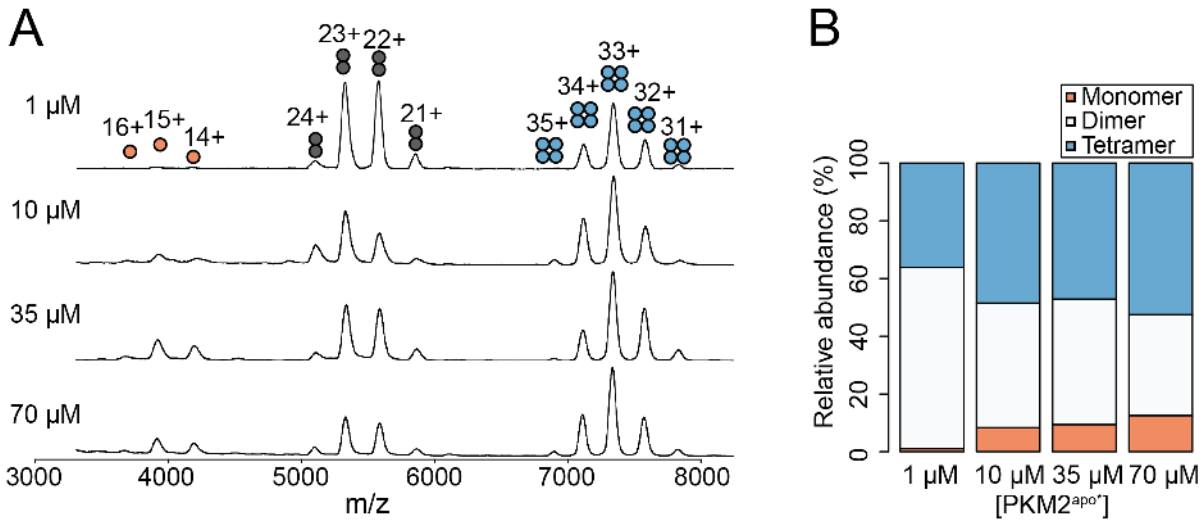

Figure 4.1: ***The concentration-dependence of PKM2 oligomerisation.*** ***(A)*** *Nano-electrospray ionisation mass spectrometry was used to acquire native spectra of PKM2^{apo*}. The protein was sprayed from a 100 mM ammonium acetate solution containing a total PKM2 concentration of 1 μM, 10 μM, 35 μM and 70 μM.* ***(B)*** *The oligomeric state distribution was quantified for each of the concentration of PKM2 and presented as stacked bars showing the relative intensities of monomers (orange), dimers (white) and tetramers (blue).*

4.2.2 PKM2 dimers form stably about the A-A' interface

PKM2 dimers might form two possible assembles about the A-A' and C-C' interfaces

The inherent two-fold symmetry of PKM2 tetramers may give rise to one of two possible dimeric assemblies about the A-A' interface (the A-A' dimer) or about the C-C' interface (the C-C' dimer) (**Fig. 4.2 A**). It was unclear whether the PKM2 dimers observed in the native spectra (**Fig. 4.1 A**) were of the A-A' or C-C' assembles, or a mixture of the two. PKM2 protomers have the same primary structure and hence, identical masses, and so it was not possible to mass-resolve the two possible species. Nevertheless, an inspection of the overall shape of *in silico* models generated of the two dimeric assemblies showed a difference in shape between the two possible dimer species (the A-A' dimer is more globular than the elongated C-C' dimer assembly; **Fig. 4.2 A**). As such, we hypothesised that the two dimer assemblies could be distinguished by their mobilities in the gas phase.

***In silico* calculations find that A-A' and C-C' dimers have distinct ion mobilities**

To this end, we used IM-MS to measure the experimental drift time of selectively activated PKM2 dimer charge states. The drift time of the ion in a defined buffer gas (t_D) is proportional to the rotationally-averaged collision cross section (CCS, Ω), which can be interpreted as the shape adopted by a given molecular ion under particular gas phase conditions[166]:

$$\Omega = \frac{\sqrt{18\pi}}{16}\frac{ze}{\sqrt{k_B T}}\left[\frac{1}{m_I}+\frac{1}{m_N}\right]^{0.5}\frac{t_D E}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} \tag{4.1}$$

where e is the elementary charge, E is the charge carried by the ion, N is the number density, L is the path length of the drift cell, T is the temperature and k_B is the Boltzmann constant.

For comparison with experimental measurements, the CCS of the 23+ A-A' and C-C' dimers, the 15+ monomer and the 33+ tetramer ions were computed from short 10 ns molecular dy-

namics simulations of A-A' and C-C' dimer *in silico* models (see Methods Section 2.2.2). The CCS was calculated from the representative clustered structures using a projection-approximation (PA) approach (^{PA}CCS), which has been used for many years to calculate rotationally-averaged CCS values for molecules[166]. PA methods, however, fail to account for long-range interactions and all the physical details of the scattering process between the ion and the buffer gas[166]. For a more explicit treatment of the scattering process in calculating the CCS, ^{PA}CCS was supplemented with additional CCS calculations using an elastic hard-sphere scattering (EHSS) approach ($^{EHSS}CCS_{He}$), in which the CCS is calculated by averaging the momentum transfer cross section which is related to the scattering angles between the incoming and departing gas atom trajectory[166].

A comparison between theoretical and experimental CCS measurements finds that dimeric PKM2 forms the A-A' assembly

The experimentally determined collision cross section for the 15+ monomer ion ($^{DT}CCS_{He}$) was (34.5 ± 9.9) nm^2, 3.3 % larger than the theoretical $^{EHSS}CCS_{He}$ (**Fig. 4.2 B**). Although the difference between the $^{EHSS}CCS_{He}$ and the experimental $^{DT}CCS_{He}$ was found to be significant (p-value < 0.001), a more pronounced structural compaction in gas-phase molecular dynamics simulations has been previously reported to manifest in smaller collision cross sections, relative to the experimental measurements[167]. The theoretically determined ^{PA}CCS was found to further underestimate the experimental $^{DT}CCS_{He}$ by 10.1 % [^{PA}CCS = (31.00 ± 1.71) nm^2] (**Fig. 4.2 B**), likely because the PA approach does not account for the effects of multiple ion-gas collisions, and hence for ions of masses greater than 2 kDa, this method will underestimate the CCS and is know to be of limited use for the study of larger biomolecules[166].

The experimentally measured $^{DT}CCS_{He}$ for the 33+ tetramer ion [(92.46 ± 2.57) nm^2] was in good agreement with the theoretically-determined CCS as determined using the EHSS-approach [$^{EHSS}CCS_{He}$ = (92.98 ± 3.41) nm^2] (**Fig. 4.2 B**). Taken together, the agreement between theoretically and experimentally determined CCS values for PKM2 monomer and

tetramer ions suggested that similar measurements of the dimer ion would have predictive power in elucidating the assembly of dimeric PKM2 in the gas phase.

The C-C' dimer was found to have a theoretical CCS of [$^{EHSS}CCS_{He}$ = (61.08 ± 6.73) nm^2], 13 % larger than the A-A' dimer [$^{EHSS}CCS_{He}$ = (53.20 ± 2.23) nm^2] (**Fig. 4.2 B**), suggesting that the two possible dimer assemblies have distinct ion mobilities. Experimental CCS measurements of the 21+ dimer ion [$^{DT}CCS_{He}$ = (52.59 ± 9.37) nm^2] were in close agreement with the theoretically-determined CCS of the A-A' dimer (**Fig. 4.2 B**). The close match between the theoretical CCS of the A-A' dimer and the experimentally-measured CCS of dimeric PKM2, suggested that PKM2 dimers form the A-A' assembly in the gas phase.

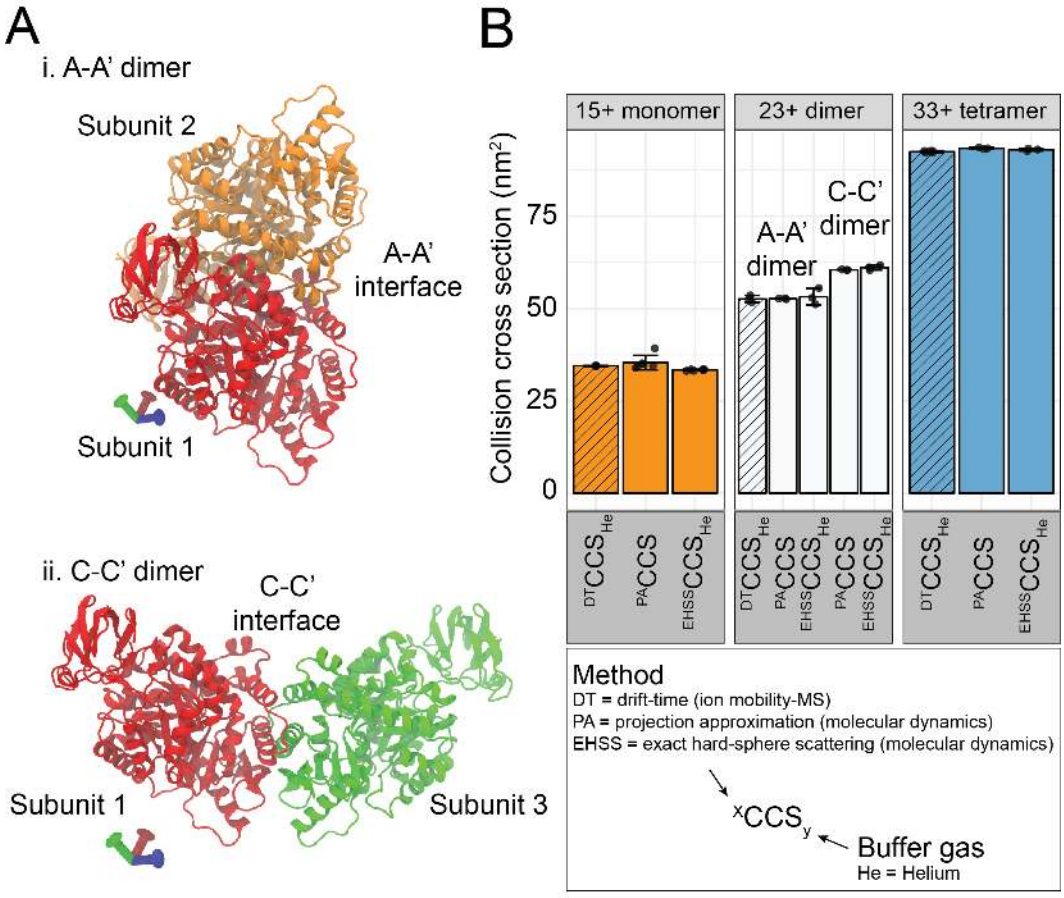

Figure 4.2: ***Ion mobility measurements reveal that PKM2 dimers are stable about the A-A' interface.*** ***(A)*** *Structural models of [i] A-A' dimers and [ii] C-C' dimers, shown in cartoon representation.* ***(B)*** *The CCS of the 15+ monomeric (orange), 23+ dimeric (white) and 33+ tetrameric (blue) $PKM2^{apo*}$ calculated experimentally by IM-MS (shaded bars) and from MD simulations of the respective ions using the Projection Approximation (^{PA}CCS) and Exact Hard-Sphere Scattering ($^{EHSS}CCS_{He}$) approaches.*

4.2.3 FBP induces PKM2 tetramerisation

Consistent with previous studies[19,48,55,60], a characterisation of enzyme kinetics of PKM2 in Chapter 3 showed that the binding of FBP increases the effective substrate affinity of PKM2 for phosphoenolpyruvate. To explore whether this functional effect of activator binding was accompanied by an induced change in the oligomeric state, we acquired native spectra of PKM2 titrated with FBP. FBP addition was found to increase the relative abundance of tetrameric charged-state species in a dose-dependent manner (**4.3 A** and **B**). Additionally, an apparent m/z shift in the tetrameric peaks and a qualitative change to the overall charged-state envelope was observed (**4.3 C**). The m/z shift was expected to result from the binding of at least three molecules of FBP to each PKM2 tetrameric species, given that preparation of protein had been previously characterised as containing $\simeq$ 23 % of the binding sites pre-occupied with co-purified FBP (Section 3.2.1).

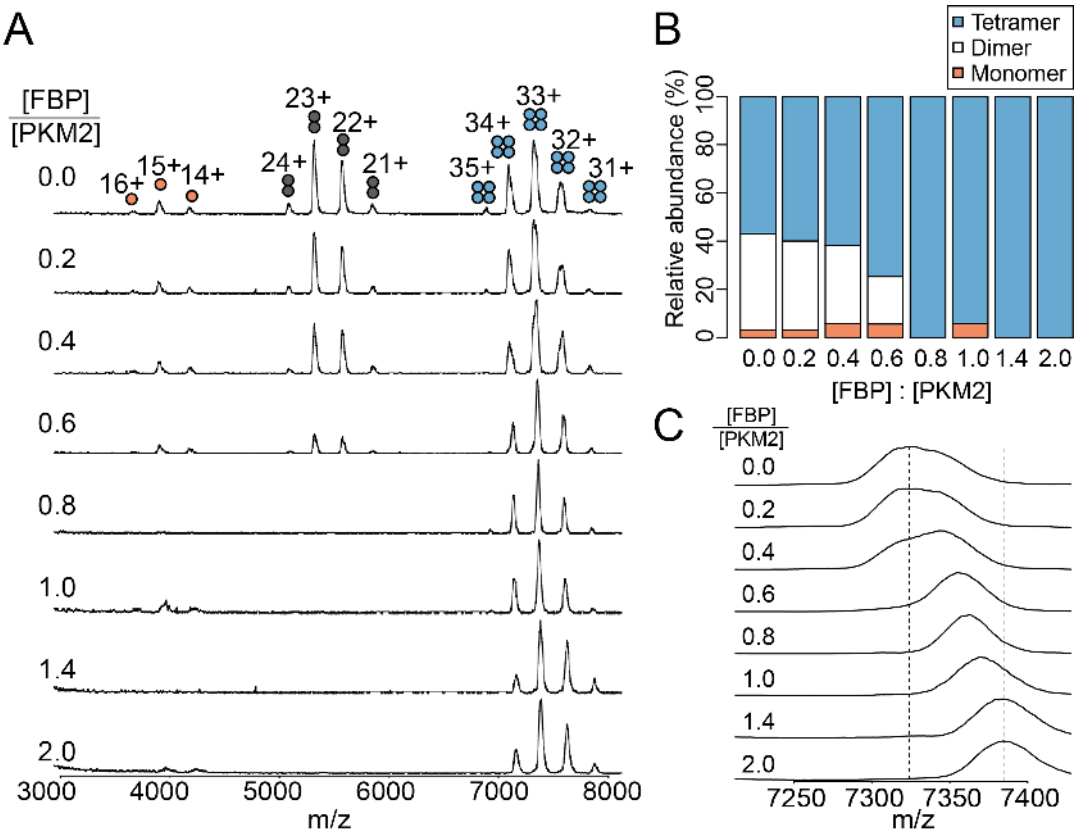

Figure 4.3: ***FBP addition induces a dose-dependent tetramerisation of PKM2.*** ***(A)*** *Native spectra were acquired for PKM2 titrated with FBP. The ratio of protein-to-ligand is indicated.* ***(B)*** *The relative intensities of oligomeric states was determined for each ratio-metric mixture of FBP-PKM2.* ***(C)*** *Native spectra of the 33+ tetrameric charge state species for each stoichiometric mixture of FBP-PKM2.*

4.2.4 PKM2^{Apo*} tetramers are formed from a mixture of apo and holo protein

In order to resolve the apparent mass shift of tetrameric PKM2 upon FBP addition, spectra of PKM2 were acquired under high cone-voltage conditions in order to remove salt adducts during the ionisation process, which would otherwise add to the overall weight of analyte ions. Mass-deconvolution of the subsequent spectra (see Methods Section 2.4.1) could thus be compared to the theoretical mass of the PKM2 from its primary sequence.

The mass-deconvolved spectrum of FBP-saturated PKM2 ($PKM2^{FBP}$) revealed a single mass peak at 234.190 kDa (**Fig. 4.4**) corresponding to the theoretical mass of tetrameric PKM2 bound to four molecules of FBP (234.273 kDa). In contrast, tetrameric PKM2 in the absence of any added FBP ($PKM2^{apo*}$), consisted of five separate peaks (**Fig. 4.4** and Table 4.2). The difference in mass between the different $PKM2^{apo*}$ mass peaks were approximately equivalent to the mass equivalents of a single molecule of FBP (340 Da), suggesting that the peaks reflected tetrameric PKM2 species bound to 0, 1, 2, 3 and 4 molecules of FBP (**Fig. 4.4**). The intensities of the five mass peaks of PKM2 in the absence of added FBP, reaffirmed the finding that considerable molar amounts of FBP are retained during the purification of PKM2 (Section 3.2.1). Additionally, these results show that FBP binding is not required for PKM2 tetramer formation. Moreover, the relative apex-intensities of the five mass peaks of $PKM2^{apo*}$ may provide information on the degree of cooperativity associated with FBP binding. For an infinitely cooperative system, where the concentration of ligand is limiting, one would expect appreciable quantities of only the *apo* and *holo* species. The distribution of peaks with an approximate ratio of 5:2:2:2:1, however, suggests that FBP binding is weakly cooperative.

Molecule	Mass (Da)
Monomeric PKM2	58218.16
Tetrameric PKM2	232872.60
D-fructose-1,6-bisphosphate	350.0
Tetrameric PKM2 + 1 FBP	233222.6
Tetrameric PKM2 + 2 FBP	233572.6
Tetrameric PKM2 + 3 FBP	233922.6
Tetrameric PKM2 + 4 FBP	234272.6

Table 4.1: Theoretically-determined masses of PKM2 and FBP

System	Conditions	Number of tetramer peaks	Mass (Da)
10 μM PKM2	High voltage	4	232880, 233220, 233580, 233930, 234290
+ 10 μM FBP	Native-like	1	235030
+ 10 μM FBP	High voltage	1	234190

Table 4.2: Maximum entropy mass estimation of tetrameric PKM2 from nESI-MS m/z spectra under native-like and high cone-voltage ionisation conditions.

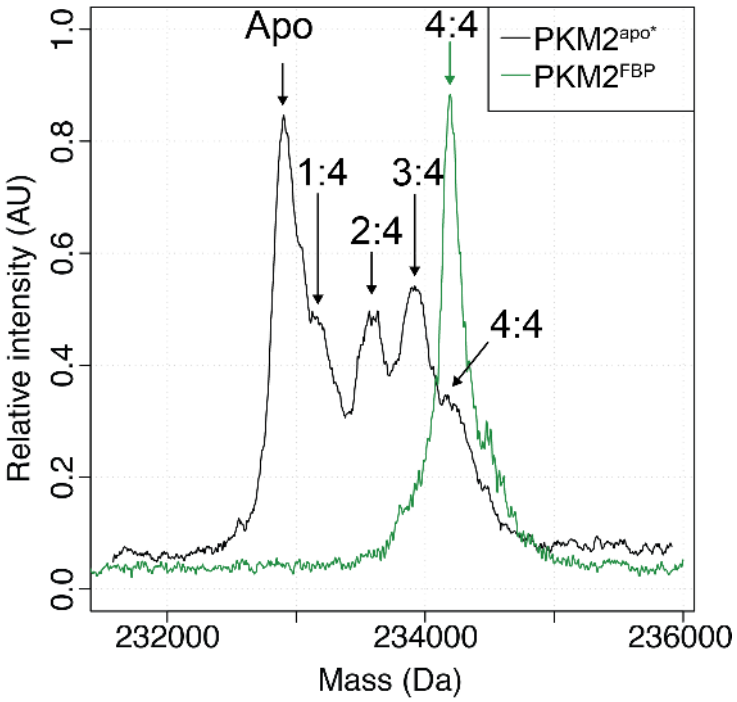

Figure 4.4: ***The mass-deconvolved spectrum of*** $PKM2^{apo*}$ ***reveals a mixture of apo and holo species.*** *Mass-deconvolved spectra were calculated from spectra of* $PKM2^{apo}$ *and* $PKM2^{FBP}$*. Normalised mass spectra of* $PKM2^{apo*}$ *(black) and* $PKM2^{FBP}$ *(green) are shown. Peaks are annotated with the corresponding FBP-PKM2 stoichiometric species, as determined from the exact mass of the measurement.*

4.2.5 FBP binding induces subtle changes to the shape of tetrameric PKM2

Previous studies have suggested that allosteric ligands induce inter-domain conformational changes within the PKM2 tetramer[22,55,62,77] and other pyruvate kinase homologues[66,72,74,75]. Therefore, we investigated the possibility that FBP binding induced conformational changes that could be detected by an altered mobility in the gas phase.

To this end, IM-MS measurements of PKM2 were acquired in the absence and in the presence of saturating amounts of FBP, from which a global CCS was calculated for each tetramer (see Methods Section 2.4.2). We found that differences in the $^{DT}CCS_{He}$ distribution, which reflects the conformational heterogeneity of proteins, suggested that FBP caused subtle conformational changes in the tetramer (**Fig. 4.5**). The negative difference in the $^{DT}CCS_{He}$ distribution between PKM2^{apo*} and PKM2^{FBP} at $^{DT}CCS_{He} > 10000$ suggested that FBP binding marginally reduces the compaction of the tetramer.

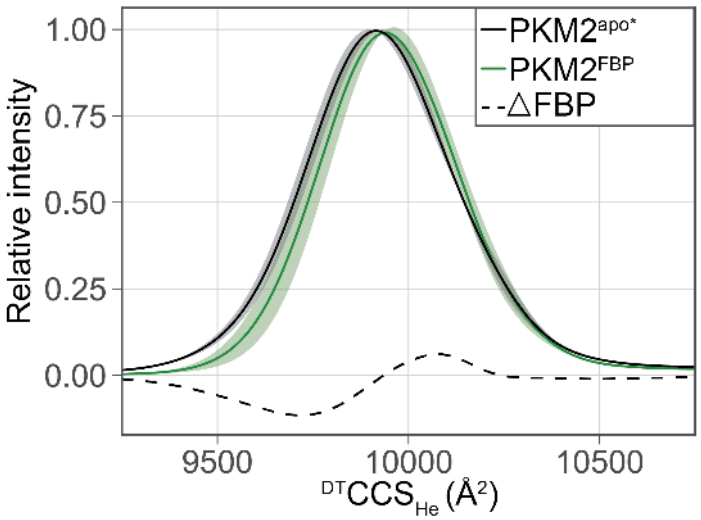

Figure 4.5: ***FBP binding changes the global collision cross section of PKM2 tetramers.*** *$^{DT}CCS_{He}$ distribution of PKM2^{apo*} and PKM2^{FBP} calculated from analyses of arrival time distribution measurements of PKM2 tetramer peaks (see Methods Section 2.4.2). Average calculated values are shown in solid lines and the standard deviations are shown as shaded ribbons. The difference spectrum between PKM2^{apo*} and PKM2^{FBP} is shown as a dashed line.*

4.3 Inhibitory amino acids have distinct effects on tetramerisation depending on whether FBP is bound

Measurements of the steady-state kinetics of PKM2 in Chapter 3 found that while FBP addition promoted high substrate affinity, simultaneous addition of phenylalanine (Phe) and FBP prevented FBP from exerting its maximally activating allosteric effect by reducing the k_{cat}. nESI-MS experiments supported the hypothesis that FBP activation involves tetramerisation of PKM2, linking activation to the oligomeric state. Competing views of amino acid inhibition of PKM2 debate the effects of Phe and Ala binding on PKM2 oligomerisation[20,22,70]. We therefore set out to determine the effects of amino acids on the oligomeric state of PKM2, and to explore whether Phe impedes FBP-induced activation of PKM2 by affecting changes to the oligomeric state of the protein.

4.3.1 Phe addition does not disrupt FBP-induced PKM2 tetramerisation

Addition either Phe to PKM2 at a protein-to-ligand ratio of 30:1, resulted in an increase in the relative abundance of dimeric PKM2 charge state species (**Fig. 2.1 A** and **B**), suggesting that Phe addition *per se* favoured the dimeric form of PKM2. The perference of Phe for dimerisation is in agreement with previous solution-phase analytical ultracentrifugation studies[70,168]. Broadening of the dimeric and tetrameric charge state peaks was observed, likely due to additional salt molecules binding to the protein upon addition of high concentrations of the ligand (**Fig. 2.1 A** and **B**). The apparent effects of Phe addition in promoting the dimeric state of PKM2, while significant in its magnitude, may have been underestimated due to the stoichiometry of binding. Considering the measured binding affinities of both ligands, calculated estimates of the fraction of 10 μM PKM2 bound upon addition of 300 μM Phe is 0.68, leaving a considerable fraction of the unbound protein. Attempts at measuring native m/z spectra of PKM2 at higher ligand concentrations were unsuccessful due to the unacceptably high salt content of the protein-ligand mixture.

Next, native spectra of PKM2 were acquired following pre-incubation with saturating concentrations of FBP. As before (Section 4.2.3), FBP addition resulted in PKM2 tetramerisation (**Fig. 4.6 A** and **B**). Addition of Phe subsequent to pre-incubation with FBP did not perturb the tetrameric state of PKM2, while FBP binding to PKM2 pre-incubated with Phe resulted in protein tetramerisation, indicating that the dominant effect of FBP on PKM2 tetramerisation was not effected by the order of ligand addition (**Fig. 4.6 A** and **B**).

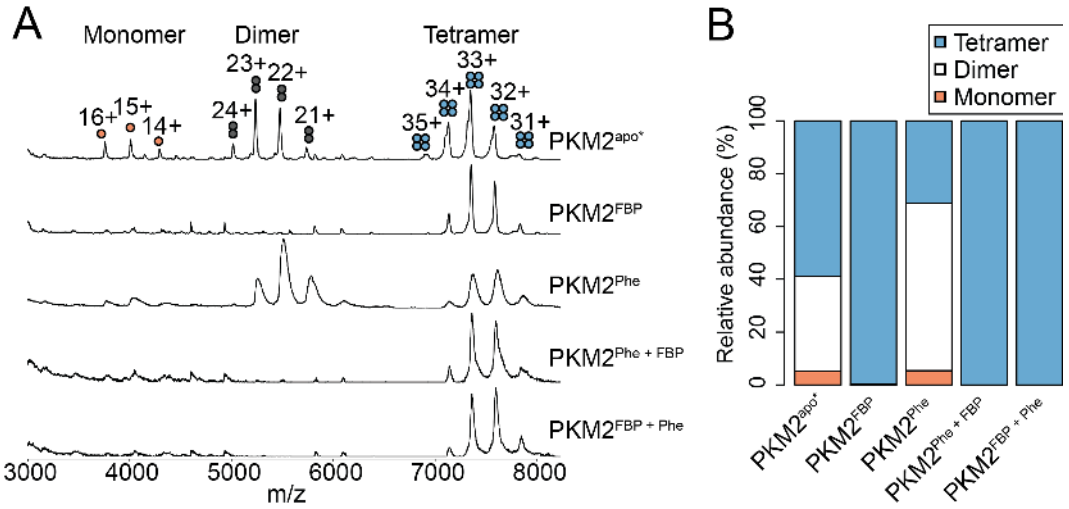

Figure 4.6: ***Phe binding does not disrupt FBP-induced PKM2 tetramerisation.*** ***(A)*** *Native spectra of 10 μM $PKM2^{apo*}$ and in the presence of 10 μM FBP ($PKM2^{FBP}$), 300 μM Phe $PKM2^{Phe}$, 300 μM Phe followed by addition of 10 μM FBP $PKM2^{Phe+FBP}$ and 10 μM FBP followed by addition of 300 μM Phe $PKM2^{FBP+Phe}$. Relative oligomeric state abundances were calculated and are shown in* ***(B)****.*

4.3.2 Phe and FBP synergistically promote PKM2 tetramerisation

It appeared from from the preceding native spectra of PKM2, that Phe had a neutral effect on the oligomeric state of FBP-bound PKM2 and that their binding was simply a passenger to FBP-induced tetramerisation. This hypothesis was found not to hold in the context of sub-stoichiometric FBP addition, where the subsequent addition of either Phe was found to act synergistically with FBP to promote PKM2 tetramerisation with slow kinetics [k_{tet} = (812.5 ± 284.6) s^{-1}] (**Fig. 4.7**). Conversely, the addition of equivalent half-stoichiometric amount of FBP in the absence of Phe, were unable to fully convert PKM2 monomers and dimers into tetramers (**Fig. 4.7**). The propensity for Phe to enhance FBP-induced tetramerisation implied a functional synergism between the activator and either amino acid inhibitors, favouring tetramer formation despite to apparent opposing effects of these ligands, individually, both on enzyme activity and oligomerisation.

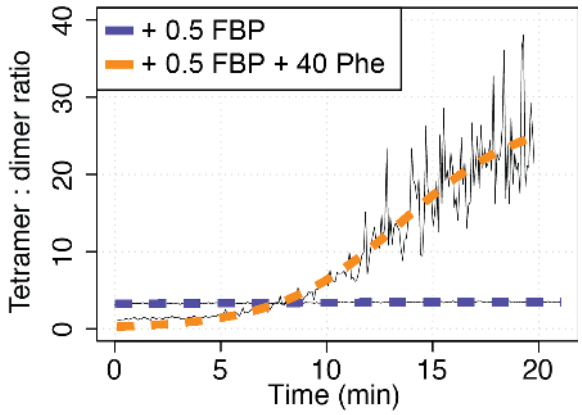

Figure 4.7: ***FBP and Phe synergistically promote PKM2 tetramerisation.*** *The time-resolved ratio of tetramer-to-dimer charge state intensities of 10 μM PKM2 following pre-incubation with either (i) 5 μM FBP, or (ii) 5 μM and 400 μM Phe. Time-resolved intensities are fit to a two-state sigmoidal model.*

4.3.3 Analytical evidence for simultaneous binding of FBP and Phe to PKM2

The failure of Phe addition to destabilise FBP-bound PKM2 tetramers suggested a dominant effect of FBP on the oligomeric state of the protein. Moreover, it was predicted that the simultaneous addition of Phe and FBP to PKM2 would result in doubly-liganded PKM2, given the the finding from ligand binding studies that the binding affinity of Phe was unaffected by FBP binding, and *vice versa*. Nevertheless, to address the possibility that under ionising conditions into the gas phase Phe does not bind to $PKM2^{FBP}$, we sought analytical evidence for simultaneous binding of Phe and FBP to PKM2.

Evidence for concurrent Phe and FBP binding to PKM2 was provided by the m/z shift produced by ligand addition to PKM2. Close inspection of the +33 tetramer charge state found that the $PKM2^{apo*}$ spectrum contained a doublet of tetramer peaks: the left portion of the peak was the fully-apo protein and the right portion resulted from ions of 1, 2, 3 and 4 FBP molecules bound to PKM2 tetramers (**Fig. 4.8 A**). Addition of saturating amounts of FBP resulted in a small m/z shift towards higher CSD values (**Fig. 4.8 A**). A further shift of the $PKM2^{FBP}$ 33+ peak was observed upon addition of Phe to FBP-bound PKM2 (**Fig. 4.8 A**).

The shift towards lower charge-states upon addition of Phe was found to be reversible when the cone voltage (CV) in the electrospray ionisation source was increased from 10 V to 100 V. Increasing the CV had the observed effect of reducing the negative charge of the $PKM2^{FBP+Phe}$ peaks and converting them to align with the $PKM2^{FBP}$ (**Fig. 4.8 A**). This was achieved by displacing Phe molecules bound to $PKM2^{FBP}$. Conversely, increasing the CV of the ESI source did not convert the +33 $PKM2^{FBP}$ ion to the 'apo-like', suggesting that the high FBP-PKM2 binding affinity stabilised the interaction from dissociation even at high ionisation energies.

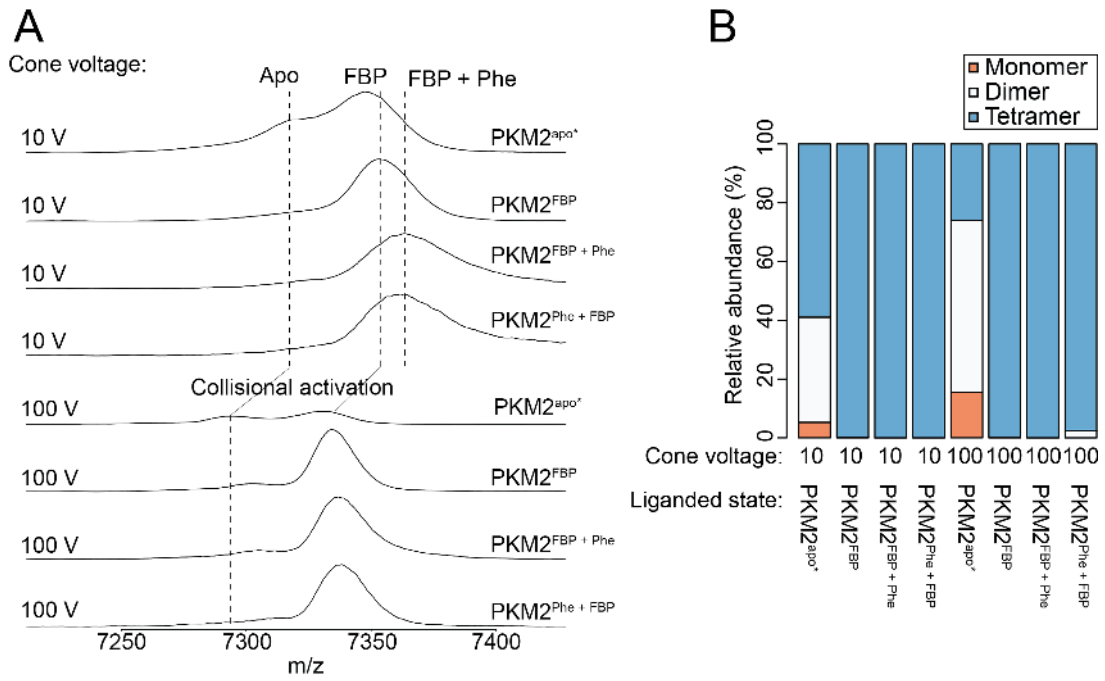

Figure 4.8: ***FBP and Phe can simultaneously bind to PKM2.*** ***(A)*** *Mass spectra of PKM2 were acquired in the absence of any added ligands ($PKM2^{apo*}$), and following the addition of stoichiometric amounts of FBP ($PKM2^{FBP}$); addition of 400 μM Phe to 10 μM PKM2 pre-incubated with FBP ($PKM2^{FBP+Phe}$); or addition of 10 μM FBP to PKM2 pre-incubated with 400 μM Phe ($PKM2^{Phe+FBP}$). Spectra were acquired with the cone-voltage of the electrospray ionisation source set at either 10 V (native-like) or 100 V (collisional activation). Positions of the m/z peaks are shown as dashed lines.* ***(B)*** *The relative abundances of monomer (orange), dimer (white) and tetramer (blue) species were quantified from the spectra in (**A**).*

4.3.4 Phe partially reverses PKM2 to an 'apo-like' conformation

Previous calculations of the CCS (Section 4.2.5) revealed subtle conformational changes in PKM2 in response to FBP binding. Given that the subsequent addition of Phe was found to reduce the $\frac{k_{cat}}{K_M}$ ratio to apo-like levels (Chapter 3) we questioned whether the apparent conformational changes might be predictive of the enzyme activity level of the protein. To this end, IM-MS measurements of PKM2 were repeated following concurrent addition of Phe and FBP. We found that simultaneous binding of Phe and FBP partially reversed the FBP-induced shift towards higher $^{DT}\mathrm{CCS}_{He}$ values (**Fig. 4.9**), suggesting that Phe might reverse the conformational changes associated with FBP-induced enzyme activation.

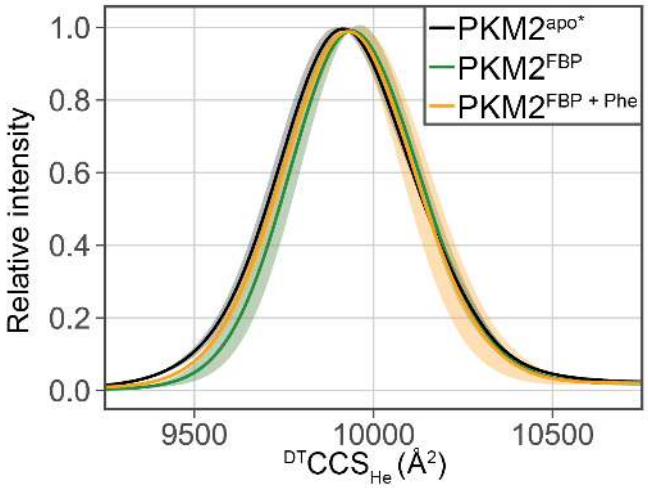

Figure 4.9: ***Phe partially reverses FBP-induced conformational changes.*** *IM-MS measurements of PKM2 in the absence of any added ligands (black) in the presence of stoichiometric amounts of FBP (green) and in the presence of FBP and 300 µM Phe (orange). A protein concentration of 10 µM was used for all measurements.*

4.4 Conclusion

Collectively, enzymology, biophysics and native mass spectrometry suggest that allosteric regulators of PKM2 exert distinct effects on PKM2 catalysis by modulating the K_M^{PEP} and the k_{cat}. Furthermore, this differential effect on alternative aspects of enzyme catalysis is proposed to be routed in the oligomeric structure of PKM2. A characterisation of amino acid and FBP binding revealed that the kinetics of ligand binding to these two pockets occcur independently. The binding of amino acids was found to occur with the same apparent dissociation constant irrespective of whether or not FBP is bound to its pocket, and *vice versa*. Concurrent binding of Phe and FBP to PKM2 was further supported by the finding that subsequent ligand addition results in further reversible shifts to the charge-state distribution of PKM2 tetramers.

However, a functional cross-talk between the two binding pockets is proposed. In support of this hypothesis we have found that the enzyme kinetics of PKM2 are distinct when both amino acids and FBP are bound simultaneously, compared to their effects *per se*. Moreover, the functional synergism between amino acid and FBP binding is manifest by enhanced kinetics PKM2 oligomerisation, despite opposing effects of Phe and Ala stabilising dimers and FBP stabilising the tetrameric state. This suggested a functional cross-talk routed in the mechanism of allosteric signal transfer from either allosteric pocket to the protomer interfaces and the active sites, possibly through cooperation between the allosteric pathways connecting these functional sites.

In order to understand the molecular basis of this functional cooperation between amino acid and FBP regulation, we sought to model the effect of these allosteric ligands on the conformational dynamics of PKM2 using molecular dynamics simulations. In addition, we developed a computational method designed to extract allosteric pathways from molecular dynamics simulations in a statistically robust manner.

Chapter 5

Molecular dynamics simulations predict hub residues involved in the allosteric transition of PKM2.

5.1 Introduction

Allosteric ligand binding was shown in the previous two Chapters to affect changes to PKM2 activity by regulating the oligomeric state and the global conformational shape of the protein. Previous investigations have suggested that human PKM2[22], and some of its orthologues[66,72,75], undergo a concerted rigid-body structural motion in response to FBP binding. In contrast, very little is known about the fine-grained molecular mechanism by which FBP and amino acids regulate PKM2; both in terms of the consequent energetic changes (either enthalpic or entropic) and the amino acid residues involved. Moreover, the structural resolution afforded by fluorescence spectroscopy and mass spectrometry measurements is limited.

An efficient strategy for probing allosteric regulation with sufficient temporal, spatial and energetic resolution, is the use of molecular dynamics (MD) simulations to model the conformational dynamics of proteins[83,169,170]. Therefore, we sought to investigate the molecular

mechanism of PKM2 allosteric regulation with a view towards identifying the specific protein residues, which are involved in propagating activating or inhibitory effects from distinct allosteric sites to the catalytic pocket, using MD simulations. In this Chapter, results are presented from a comprehensive study of the energetics and dynamics of monomeric and tetrameric PKM2 using computer simulations. Additionally, a novel computational method *AlloHubMat* is described and applied towards extracting predicted allosteric hub residues in PKM2.

5.2 Molecular dynamics simulations of PKM2 capture ligand-induced conformational changes

To explore possible protein backbone conformational changes, elicited by allosteric ligand binding to PKM2, MD simulations of PKM2 were performed in the absence and in the presence of several allosteric ligands. Given that PKM2 exists as an equilibrium of oligomeric states (Section 4.2.1), simulations of both monomeric and tetrameric protein were performed to investigate differences in the conformational dynamics between the two oligomers. Monomeric PKM2 was simulated in the apo-form (mPKM2^{apo}), bound to the endogenous inhibitor Phe (mPKM2^{Phe}), the endogenous activator FBP (mPKM2^{FBP}) and the exogenous small-molecule activator Tepp-46 (mPKM2^{Tepp}). Similarly, to investigate the dynamics involved in the ligand-induced conformational transition of PKM2 (Section 4.2.5), simulations of tetrameric PKM2 were seeded from protein crystal structures in the apo-form (tPKM2^{apo}), bound to FBP (tPKM2^{FBP}), concurrently bound to serine and FBP (tPKM2$^{FBP+Ser}$) and bound to phenylalanine and FBP (tPKM2$^{FBP+Phe}$).

5.2.1 An integration time step of 2 fs appropriately conserves the energy in monomeric PKM2 simulations

Prior to production runs in the NPT ensemble, short 40 fs simulations of mPKM2^{apo} were performed in the microcanonical (NVE) ensemble to evaluate the conservation of total energy of the system with integration time step sizes of 2 fs, 3 fs and 4 fs.The total energy fluctuations of the system were compared to those of the potential energy and the kinetic energy:

$$\Delta E = \sqrt{(E - \langle E \rangle)^2} \tag{5.1}$$

where E is either the total, kinetic or potential energies of the system.

Increasing the integration time step of the short NVE simulations resulted in a considerable increase in the amplitude of the fluctuations of the potential and kinetic energies (**Fig.**

5.1). To reduce errors propagated from integrating the equations of motion, fluctuations of the total energy should be less than one-fifth of the fluctuations of the kinetic or the potential energies[171]. This criteria was not fulfilled for simulations using time steps larger than 2 fs (Table 5.1), and so an integration time step of 2 fs was used in all subsequent simulations of monomeric and tetrameric PKM2.

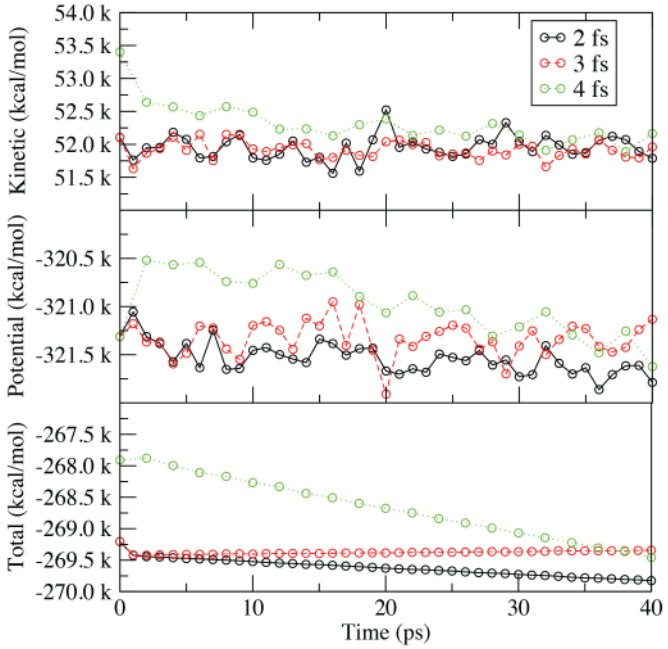

Figure 5.1: ***Energy fluctuations calculated for different integration time-steps.*** *Simulations with constant energy (NVE) were performed to evaluate how well the total ennergy was conserved for 4 fs simulations of mPKM2^{apo}, using integration time steps sizes of 2 fs (black), 3 fs (red) and 4 fs (green).*

Time step	$\frac{\Delta E_{total}}{\Delta E_{kinetic}}$	$\frac{\Delta E_{total}}{\Delta E_{potential}}$
2.0	0.2303	0.2317
3.0	0.9811	0.8420
4.0	1.3500	1.5445

Table 5.1: ***Single precision NVE simulation without thermostat or barostat coupling.*** *Quantification of the energy fluctuations in NVE simulations of mPKM2^{apo}, as shown in* ***Fig. 5.1****.*

System	Oligomer	Ligand	Replicas	Length (ns)
$mPKM2^{apo}$	Monomer	Apo	3	300
$mPKM2^{FBP}$	Monomer	Fructose 1,6-bisphosphate	5	300
$mPKM2^{Phe}$	Monomer	L-phenylalanine	3	300
$mPKM2^{Tepp}$	Monomer	Tepp-46	3	300
$tPKM2^{apo}$	Tetramer	Apo	5	400
$tPKM2^{FBP}$	Tetramer	Fructose 1,6-bisphosphate	5	400
$tPKM2^{FBP+Ser}$	Tetramer	Fructose 1,6-bisphosphate and L-serine	3	400
$tPKM2^{FBP+Ser}$	Tetramer	Fructose 1,6-bisphosphate and L-phenylalanine	3	400

Table 5.2: ***Summary of MD simulations of monomeric and tetrameric PKM2.*** *A tabular summary of explicit solvent molecular dynamics simulations performed of monomeric and tetrameric PKM2. In addition to the five replicas of 300 ns simulation of $mPKM2^{FBP}$, a single replica was continued for an additional 200 ns.*

5.2.2 FBP binding causes PKM2 monomers to sample two distinct conformational states

To investigate the mechanical response of PKM2 upon FBP and Phe binding we performed MD simulations of mPKM2^{apo}, mPKM2^{FBP} and mPKM2^{Phe} at a constant temperature of 300 K and a pressure of 1 atm. To simplify the high-dimensionality of the trajectories and to study the molecular determinants related to the binding Phe or FBP, principal component analyses were performed of the positional coordinates of the three simulations. After removing roto-translational degrees of freedom, the covariance matrix (σ) of the atomic positional fluctuations was calculated with the elements:

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{5.2}$$

where $\{x_1, x_2, x_3, ..., x_{3N}\}$ are the Cartesian coordinates of the protein. The covariance matrix was then trivially expressed in mass-weighted coordinates:

$$\sigma' = M\sigma \tag{5.3}$$

where M is the mass matrix of the protein. From the mass-weighted covariance matrix (σ'), eigenvalues (λ) and eigenvectors (x) were determined through a linear transformation:

$$0 = (\sigma' - I\lambda)x \tag{5.4}$$

where I is the unit matrix. The eigenvalue problem in Equ. 5.4 was solved for the eigenvalues and the eigenvectors of the system.

Simulated trajectories of mPKM2^{FBP} were found to sample two discrete conformational states in eigenvector space over the course of a 500 ns simulation (**Fig. 5.2 A**). A k-means clustering of the PCA plot found that two clusters [i] and [ii] explained all of the point variability of the data set. Transition between clusters [i] and [ii] was dominated by the movement of the

B-domain into the closed conformation over the substrate-binding pocket, and a change to the N-terminal helix-loop-helix (HLH) into an alternative, stable conformation (**Fig. 5.2 A**). The nature of the B-domain motion led to the annotation of cluster [i] as the *open* conformation and cluster [ii] as the *closed* conformation.

MD simulations of mPKM2^{apo}, mPKM2^{Phe} and mPKM2$^{Tepp-46}$ were similarly analysed for positional variance about the first two eigenvectors by transforming their positional coordinates into eigenvector space (**Fig. 5.2 B**). In order to project all four monomer trajectories into the same eigenvector space, the same rotation matrix was applied to each coordinate system. Analyses of mPKM2^{apo} and mPKM2^{Phe} trajectories were confined to cluster [i], thus sampling the *open* conformation, whereas $mPKM2^{Tepp-46}$ was found to sample the closed conformation.

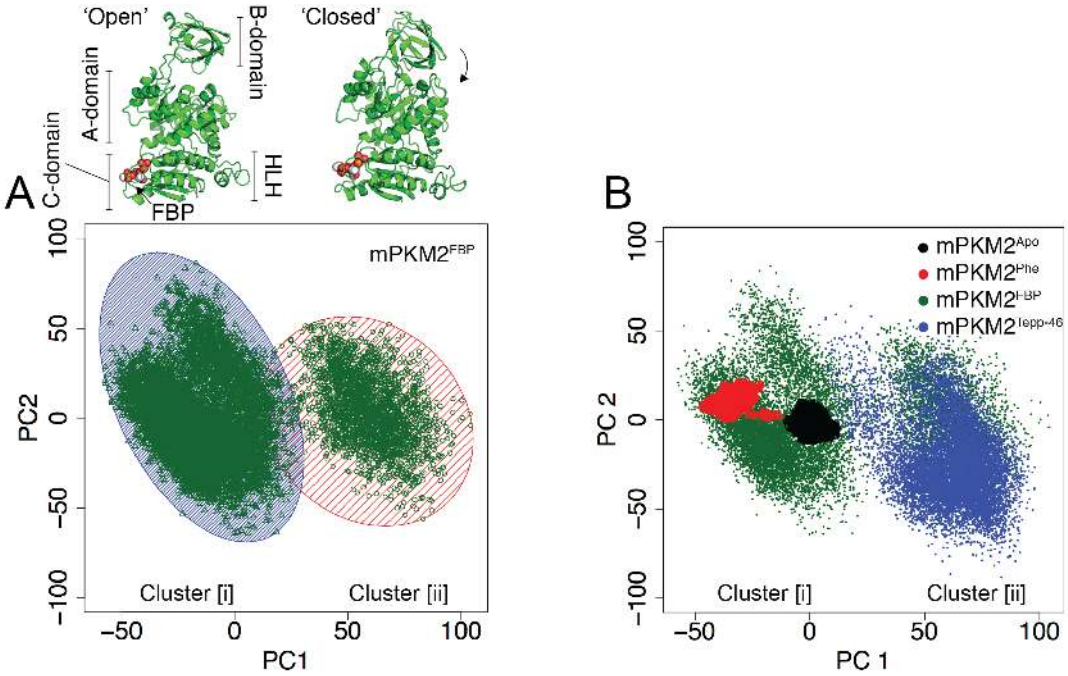

Figure 5.2: ***Conformational plasticity and essential dynamics of monomeric PKM2. (A)*** *PCA projection of a 500 ns MD simulation of PKM2 bound to FBP. Two clusters were found to explain 100 % point variability of the data set. Single most dominant conformations of $mPKM2^{FBP}$ were extract from clusters [i] and [ii] and are shown in cartoon representation above. The first eigenvector accounted for 37.5 % of the total variance and the second eigenvector accounted for 13.1 % of the total variance.* ***(B)*** *Superimposition of eigenvalues from $mPKM2^{apo}$ (black), $mPKM2^{Phe}$ (red) and $mPKM2^{Tepp-46}$ (blue) onto the first two eigenvectors determined from an eigenvalue decomposition of the Cα coordinates of $mPKM2^{FBP}$ (green).*

The *open* to *closed* transition of the B-domain in the first 300 ns of $mPKM2^{FBP}$ simulations was observed to occur as a single two-state transition, and a re-opening of the B-domain was not observed within this time scale. To investigate the conformational equilibrium of the B-domain state transition, MD simulations of $mPKM2^{FBP}$ were extended for a further 200 ns. The resulting B-domain cap dynamics was quantified by measuring the distance between the centre of mass of the A- and B-domains. Extended MD simulations of $mPKM2^{FBP}$ found a reversal between energy minima about the *open* and *closed* conformations, with an intermediate *semi-closed* conformation additionally detected (**Fig. 5.3 A** and **B**).

Taken together, these data suggested that the reversible closure of the B-domain in monomeric PKM2 was dependent on FBP binding, within the simulated time scales, and that this is a feature of PKM2 activation. Consistent with this interpretation, mPKM2^{Apo} and mPKM2^{Phe} were found to same the open conformation, whereas $mPKM2^{Tepp}$ sampled to closed conformation (**Fig. 5.3 C-E**). It was therefore hypothesised that B-domain closure contributes to enzyme activation.

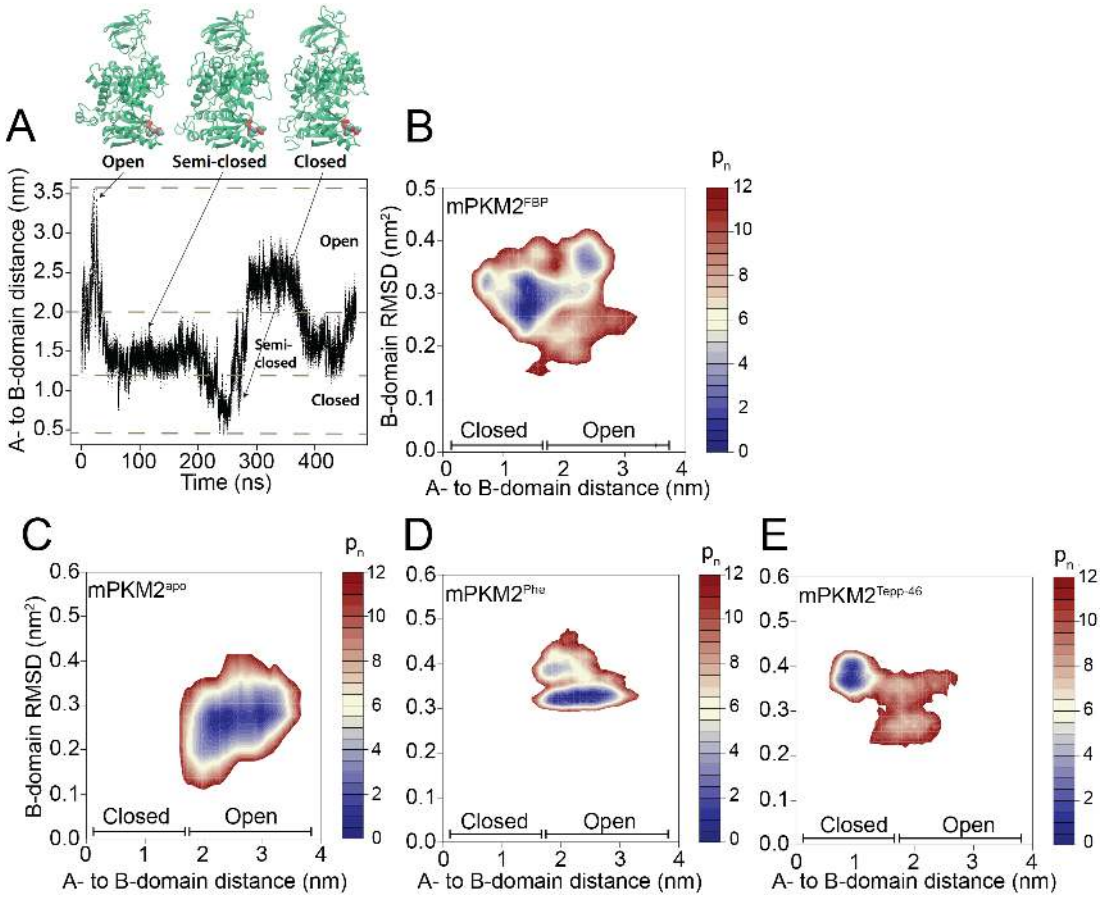

Figure 5.3: ***Conformational equilibrium of FBP-induced cap closure from a 500 ns MD simulation.*** ***(A)*** *The distance computed between the centre of mass for the A-domain and the centre of mass of the B-domain over the course of an MD simulation of* $mPKM2^{FBP}$*. Surface distribution of states plot, with Boltzmann averages of states calculated by B-domain cap distance as a function of the RMSD of the B-domain for* ***(B)*** $mPKM2^{FBP}$*,* ***(C)*** $mPKM2^{Apo}$*,* ***(D)*** $mPKM2^{Phe}$ *and* ***(E)*** $mPKM2^{Tepp}$*.*

5.2.3 B-domain closure traps highly resident water molecules in the active site

previous studies have found that a water molecule is required as part of the catalytic mechanism of PKM2 to protonate the enolate intermediate[150]. This is supported by crystal structures of PKM2 showing a cluster of water molecules proximal to the active site residues T328 and S362[23,55]. In light of the observation of ligand-dependent B-domain dynamics affecting the solvent exposure of the catalytic pocket, we postulated that the closure of the B-domain would affect the solvent dynamics in the catalytic pocket and that this may play a role in catalysis.

An analysis of water density maps[172] calculated from representative structures from MD simulations of $mPKM2^{FBP}$ in the *open* and the *closed* conformations, found that the active site pocket contained an increased number of highly resident water molecules when the B-domain cap was closed (**Fig. 5.4**). Moreover, a cluster of resident water molecules were positioned proximal to T382 and S362 in the closed state. This would suggest that B-domain closure contributes to catalysis by trapping necessary water molecules, proximal to the substrate binding pocket.

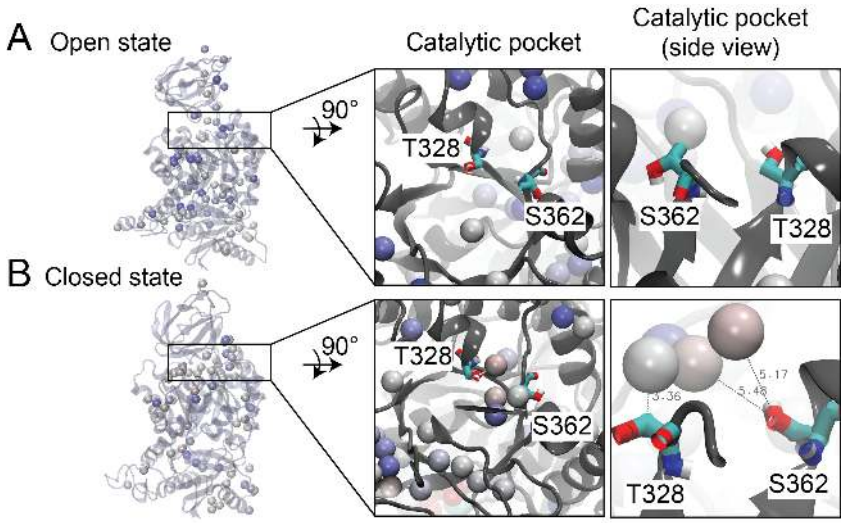

Figure 5.4: ***Water residence time at the active site of mPKM2-FBP.*** *Water density was calculated on a 0.5 Å-spaced grid and averaged over snapshots exracted from the mPKM2-FBP simulation every 0.1 ps for the* ***(A)*** *open and* ***(B)*** *closed conformations of the B-domain cap. The hydration sites (spheres) are coloured from white to blue according to increased water density.*

5.2.4 B-domain dynamics is correlated with structural changes to catalytic residues in the active site

To further explore whether B-domain closure was accompanied by other features of PKM2 activation, we investigated the side-chain dynamics at the active site of the protein. A superimposition of PKM2 active site residues from crystal structures deposited in the protein data-bank, found that the catalytic residue H78 adopts an altered side-chain conformation when the protein is bound to allosteric activators FBP and Tepp-46 (PDB ID: 3u2z) and L-serine and FBP (PDB ID: 4b2d), compared to crystal structures of apo PKM2 (PDB ID: 3bjt) and PKM2 bound to the inhibitor phenylalanine (PDB ID: 4fxj) (**Fig. 5.5 A**). H78 has been implicated the phospho-transfer reaction of PKM2 as a proton-donor and acceptor in the catalytic cycle[150]. In the activator-bound structures, the $N_{\delta 1}$ group was found to be positioned towards the active site, and was found positioned away from the active site in apo- and Phe-bound structures (**Fig. 5.5 A**). Consistent with this change following allosteric activator binding, the conformation of H78 was proposed to provide a potential read-out for whether PKM2 was in the active or the inactive states.

To investigate possible changes to the orientation of H78 in the MD simulations of PKM2, the χ_2 torsion angle of H78 was measured in simulations of mPKM2^{FBP} and mPKM2^{apo}. We found that the closure of the B-domain was accompanied by the catalytic H78 adopting a χ_2 torsion angle of between 100° and 110°, in agreement with the χ_2 torsion angles of the crystal structures of activator-bound PKM2 (108.2° for PKM2$^{FBP+Ser}$ and 103.0° for PKM2$^{FBP+Tepp}$; **Fig. 5.5 A**). Conversely simulations of mPKM2^{apo} and mPKM2^{Phe} displayed variable H78 χ_2 torsion angles approximately equivalent to that the inhibited crystal structures ($\simeq -70°$; **Fig. 5.5 A**). This further suggested that the *open* conformation corresponds to the inactive state of PKM2 and that FBP-induced closure of the B-domain contributed to the transition to the active state.

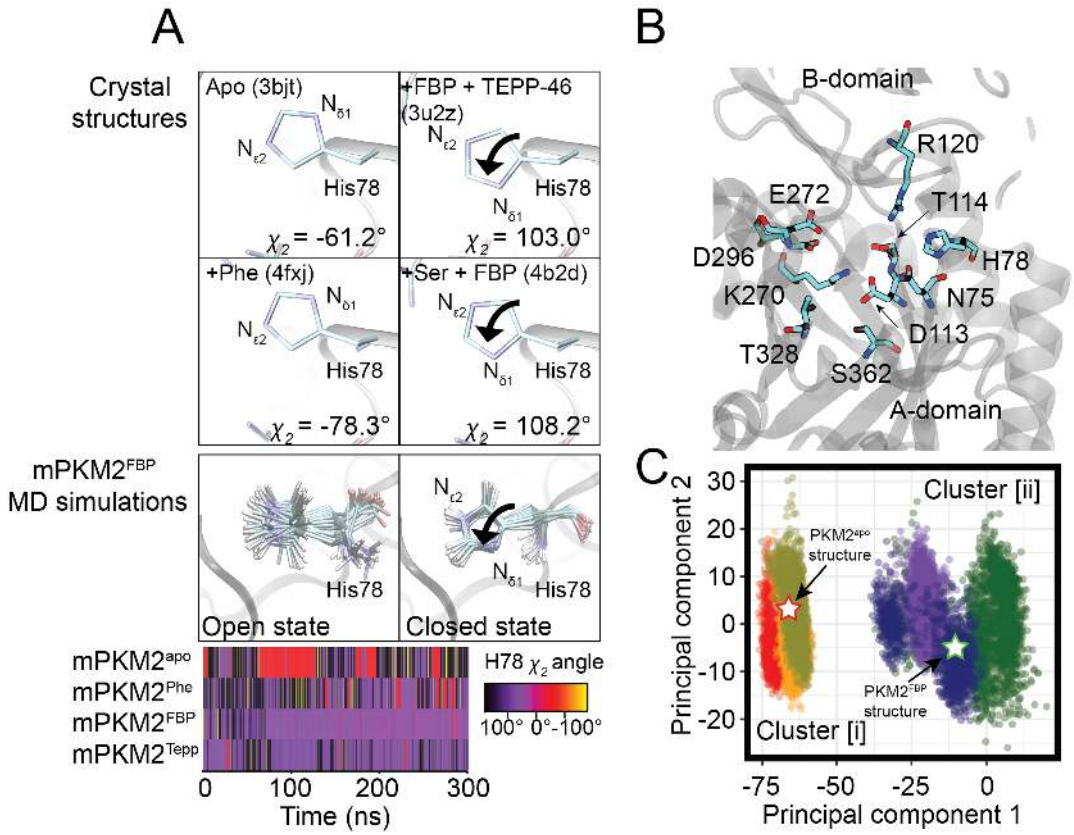

Figure 5.5: ***Active site residue H78 adopts a distinct side-chain conformation depending on the liganded state of the protein.*** ***(A)*** *The active site residue histidine 78 (H78) shown in stick representation for crystal structures of apo- FBP and activator-, phenylalanine-, and serine and FBP-bound PKM2. Below, 18 evenly spaced snap-shots of the conformation of H78 are shown for a MD simulated trajectory of $mPKM2^{FBP}$ in the open and following the transition to the closed states. In the open state [Cluster (i) in Fig. 5.2 A], the side chain of H78 is flexible and switches into a conformation seen in that of the active crystal structures following the transition of the simulation into cluster ii [Cluster (ii) in Fig. 5.2 B]. The time evolution of the dihedral angle of H78 along MD simulations of $mPKM2^{apo}$, $mPKM2^{Phe}$, $mPKM2^{FBP}$ and $mPKM2^{Tepp-46}$, colour-coded according to the scale on the right.* ***(B)*** *The PKM2 catalytic pocket, annotated to show the active site residues.* ***(C)*** *Cα active site corrdinates of the $mPKM2^{apo}$ active site (gold), $mPKM2^{Phe}$ (red), $tPKM2^{apo}$ (orange), $mPKM2^{FBP}$ (green), $mPKM2^{Tepp-46}$ (blue) and $tPKM2^{FBP}$ (purple) are projected onto the first two eigenvectors of $mPKM2^{FBP}$. This figure is partially adapted from[77], which is made available under a Creative Commons Attribution license (CC-BY).*

While the orientation of H78 correlated with the activity status of the crystal structures, we could not exclude the possibility that the observed flipping of the imidazole side chain of H78 was an artefact of the structure-refinement process. Therefore, a PCA analysis of the positional coordinates was performed for all active site residues (**Fig. 5.5 B**) in simulations of mPKM2^{apo}, mPKM2^{Phe}, mPKM2^{Tepp}, mPKM2^{FBP}, tPKM2^{apo} and tPKM2^{FBP}. This analysis found that the simulations could be separated into two clusters. The first cluster [i] contained coordinates of mPKM2^{apo}, mPKM2^{Phe} and tPKM2^{apo}, while the second cluster [ii] contained coordinates of mPKM2^{FBP}, mPKM2^{Tepp} and tPKM2^{FBP} (**Fig. 5.5 C**). Moreover, crystal structures of PKM2^{apo} (PDB ID: 3bjt) and PKM2^{FBP} (PDB ID: 3u2z) localized to the first and second clusters, respectively. Together, this suggested that side-chain active site changes were involved in the inactive-active transition. Nevertheless, the partition of *active* and *inactive* simulations into the two clusters observed in the PCA analysis of active site residues persisted for the entirety of the simulations and transitions between clusters [i] and [ii] is not observed for simulated trajectories of mPKM2^{FBP}. This suggested that an active-to-inactive transition, upon B-domain closure, does not completely describe the mechanism of enzyme activation.

5.2.5 Comparative dynamics between tetrameric and monomeric PKM2

Investigation of the oligomeric state of PKM2 by native mass spectrometry in Chapter 4 found that FBP binding results in tetramerisation. Moreover, additional experiments found that Phe and Ser compete for binding to modulate PKM2 activity, in the context of constitutive FBP binding.

There are two distinct dimer interfaces within the PKM2 tetramer; the A-A' interface between the A- and C-domains of chains 1 and 2 and chains 3 and 4, and the C-C' domains between the C-domains of chains 1 and 3, and chains 2 and 4. The C-C' domain of PKM2 is formed of adjacent C-domains aligning in a 'tail-to-tail' fashion and interacts in a four-helical bundle. Particularly prominent to the C-C' domain interface is the protrusion of Lys-421 across the interface and through the loop between residues 399 to 407 on the adjacent monomer. This loop into which Lys-421 extends has a negative electrostatic potential and thus forms the basis for several charged interactions with Glu-409 and Try-443. In contrast, the A-A' interface largely encompasses several contacts between adjacent A-domains of chains 1 and 3 and chains 2 and 4, with additional bonds formed between the B-domains and N-terminal helix-loo-helix domains of the adjacent chains. Notably, helix 11 (residues 341-353) protrude into a lipophilic pocket within the first 34 residues of the N-terminal helix-loop-helix domain and the top half of helix 12 (residues 368-378) of the opposing monomer. The apparent dissociation constant of PKM2 hetero-oligomerisation measured using MST in Chatper 3 (see **Table 3.1**) was found to be 0.9 μM and the concentration of PKM2 in three cultured cell lines was approximately 2 μM. Given that the apparent dissociation constant of PKM2 oligomerisation is similar to its cellular concentration, PKM2 may undergo reversible oligomerisation in a cellular context. The monomer-dimer-tetramer equilibrium subsequently described in Chapter 4 is likely perturbed in cells by ligands and other interacting partners. Nevertheless, a robust description of the the cellular nature of the oligomeric equilibrium of PKM2 would necessitate further *in situ* experimentation. To provide a more detailed physico-chemical model of PKM2 dynamics,

tetrameric PKM2 was simulated in the apo-form (tPKM2^{apo}), bound to FBP (tPKM2^{FBP}), bound to FBP and Phe (tPKM2$^{FBP+Phe}$) and bound to FBP and Ser (tPKM2$^{FBP+Ser}$).

In contrast to MD simulations of monomeric PKM2, analyses of tetrameric PKM2 MD trajectories did not show the same FBP-dependent lateral B-domain closure over the active site, as for mPKM2^{FBP}. Rather, an inspection of the simulated trajectories of tPKM2^{FBP} found that a network of inter-protomeric charge-charge interactions was established between R342 and two aspartate residues (D178 and D179) on a flexible loop of the B-domain on the adjacent protomer. This charged interaction was observed to form after approximately 10 ns of the simulation resulting in a *twisting* of the B-domain cap (**Fig. 5.6 A**). The positioning of the R342 side-chain was such that it blocked the B-domain cap from closing laterally over the active site, by flipping into the far side of the active site pocket (**Fig. 5.6 B**). In contrast, while the position of R342 and the other residues at the A-A' interface is largely similar between the crystal structures of tPKM2^{apo} and tPKM2^{FBP} (**Fig. 5.6 A**), an inspection of simulated trajectories of tPKM2^{apo} found that R342 flipped away from the neighbouring active site pocket and the B-domain of the opposite protomer (**Fig. 5.6 B**).

The distance between R342 and D179 was quantified from MD trajectories of tPKM2^{apo}, tPKM2^{FBP}, tPKM2$^{FBP+Ser}$ and tPKM2$^{FBP+Phe}$ (**Fig. 5.6 C**). The formation of charged interactions about the interface of tPKM2^{FBP} was reflected in the short inter-atomic distance between the R342 guanidino group and the acidic side-chain of D177, whereas MD trajectories of tPKM2^{apo} showed distances out of the range for charged-charged interactions between these two residues (**Fig. 5.6 C [i] and [ii]**). A similar molecular behaviour to $tPKM2^{FBP}$ was observed in simulated trajectories of tPKM2$^{FBP+Ser}$ in the proximity between R342 and D177, resulting from a lateral twisting of the B-domain cap (**Fig. 5.6 C [iii]**). Conversely, the A-A' interface of tPKM2$^{FBP+Phe}$ was less compact, owing to a greater distance between R342 and D177, in a manner similar to the dynamics of tPKM2^{apo} (**Fig. 5.6 C [iv]**).

Taken together, an analysis of MD simulations of tetrameric PKM2 found that the conformational dynamics of the B-domain is different to that of the monomer due to steric hindrance of lateral B-domain movement by interface residues. Nevertheless, charged interactions between R342 and D177 at the A-A' interface appeared when the tetramer was bound to allosteric activators FBP and concurrently to FBP and Ser, resulting in twisting of the B-domain and an observed structural tightening. Conversely, the apo- and FBP and Phe-bound tetramers did not show this behaviour, suggesting that ligand-dependent conformational changes may accompany PKM2 regulation.

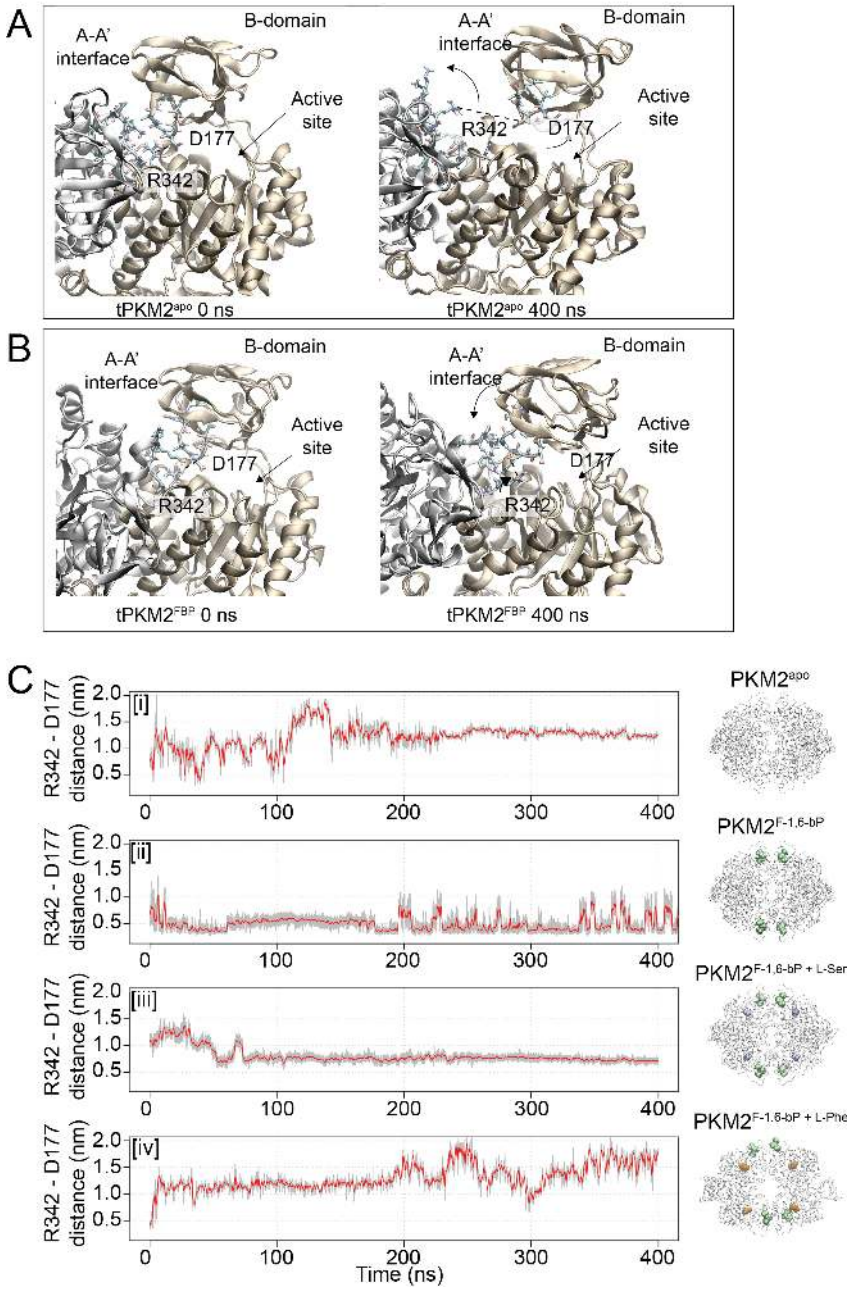

Figure 5.6: ***Charged-charged interactions at the A-A' interface prevent lateral B-domain closure in tetrameric PKM2.*** ***(A)*** *Structures of $tPKM2^{apo}$ at 0 ns and 400 ns show that R342 and D177 flip away from the A-A' interface, breaking a crucial interface charge-charge interaction.* ***(B)*** *Structures of $tPKM2^{FBP}$ at 0 ns and at 400 ns show a tightening at the A-A' interface.* ***(C)*** *The distance between R342 and D177 about the A-A' interface is quantified for [i] $tPKM2^{apo}$, [ii] $tPKM2^{FBP}$, [iii] $tPKM2^{FBP+Ser}$ and $tPKM2^{FBP+Phe}$.*

5.2.6 The configurational entropy of PKM2 does not change upon allosteric ligand binding

Entropy-driven allosteric regulation has been described in a number of proteins[110,111,173–175] where, in the absence of large-scale structural changes, allosteric ligand binding modulates the amplitude of thermal fluctuations by altering the local effective elastic modulus of the protein[86]. The configurational entropy of a macromolecular system can be calculated using a formalism first proposed by Juergen Schlitter (1993)[176], using the covariance matrix of atom-positional fluctuations.

The derivation of the Schlitter entropy for macromolecular systems is based on a quantum-mechanical treatment of a system with a of a one-dimensional degree of freedom x with states n and energies of states given by ϵ_n. The canonical partition function is given by:

$$Z = \sum_n exp\left(-\frac{\epsilon_n}{k_B T}\right) \tag{5.5}$$

where k_B is the Boltzmann constant at temperature T. The entropy of this system can be expressed by:

$$S = -k_B \sum_n p_n \, ln \, p_n \tag{5.6}$$

where p_n is the probability of finding the system in a given state, given by:

$$p_n = \frac{exp\left(\frac{\epsilon_n}{k_B T}\right)}{Z} \tag{5.7}$$

A single harmonic oscillator demands that the energy of a state is proportional to its variance ($\epsilon_n \simeq \langle n|x^2|n\rangle$), and the complete entropy of a single harmonic oscillator is given by:

$$S = \frac{k_B \alpha}{e^\alpha - 1} - k_B \, ln \, [1 - e^{-\alpha}] \tag{5.8}$$

where $\alpha = \hbar\omega\beta$, $\hbar = \frac{h}{2\pi}$, h is Plank's constant and ω is the frequency of the oscillator, which depends on the quantum-mechanically defined variance $\langle x_2 \rangle$. For a classical system the equipartition function provides the link between the quantum-mechanically defined variance and that defined in classical mechanics. Therefore, the configurational entropy of a classical system can be given by:

$$S' = \frac{1}{2} k_B \, ln \left(1 + \frac{e^2}{\alpha^2} M \sigma'_{ij}\right) \tag{5.9}$$

where σ'_{ij} is the mass-weighted covariance matrix as previously defined in Equ. 5.2.

The approximative configurational entropy was calculated for all simulations of tetrameric PKM2 using Equ. 5.9, in order to determine whether allosteric ligands might induce an entropic change to the PKM2 structure. No significant differences in the converged configurational entropy calculated from simulated trajectories of tPKM2^{apo}, tPKM2^{FBP}, tPKM2$^{FBP+Ser}$ and tPKM2$^{FBP+Phe}$ (**Fig. 5.7**), negating the likelihood of a purely entropy-driven allosteric mechanism of PKM2 regulation.

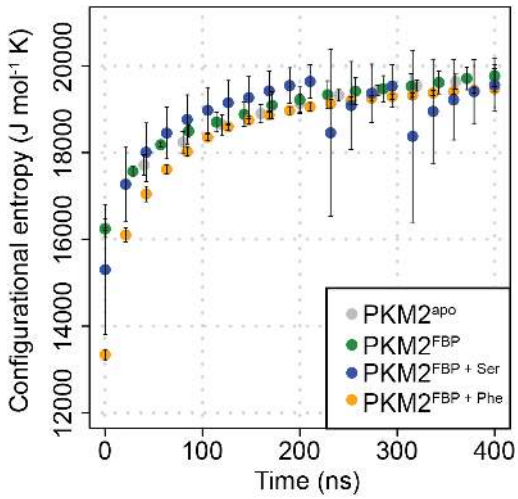

Figure 5.7: ***Allosteric ligands do not change the configurational entropy of PKM2.*** *The time-dependent configurational entropy was computed for simulated trajectories of $tPKM2^{apo}$ (grey), $tPKM2^{FBP}$ (green), $tPKM2^{FBP+Ser}$ (blue) and $tPKM2^{FBP+Phe}$ (orange).*

5.3 Identification of PKM2 allosteric hub residues using a novel software AlloHubMat

Ligand-induced changes to the collision cross section of PKM2 (Section 4.2.5) together with evidence from MD simulations of B-domain closure upon allosteric activator binding (Section 5.2.5) suggested that backbone motions within tetrameric PKM2 contribute towards the allosteric transition of the protein. In the absence of changes to the configurational entropy of PKM2 upon ligand binding (Section 5.2.6), we hypothesised that FBP binding elicits a network of correlated motions within the backbone of PKM2, and that these concerted motions form the basis of a network of residues which connect the FBP binding pocket to the active site.

We set out to quantify the network of correlated motions in MD simulations of PKM2 as a means for identifying residues involved in the allosteric mechanism. Particularly successful has been the the use of structural fragment analysis methods[177], and in particular the GSAtools developed by Pandini *et al.* (2013)[124], which were applied towards identifing distally correlated motions in the backbone of proteins in order to elucidate allosteric pathways driven by local conformational switches[123,125,128,130]. The GSAtools method uses information theory to compute the normalised mutual information (nMI) between each fragment-encoded position in the protein (Section 1.3.3). Given that each element in the nMI matrix has a time component, the network of correlated motions is likely to change over the simulated time. In this context, the time-evolution of correlated motions derived from an information theoretical treatment of the protein structure has never been explored. Moreover, no methods had been previously developed to allow for the comparison of nMI matrices extracted from multiple replicate MD trajectories.

We therefore developed a novel computational framework, named AlloHubMat (**Allo**steric **Hub** prediction using **Mat**rices that capture allosteric coupling), to predict allosteric hub fragments from the network of dynamic correlated motions, based on explicitly identified con-

Function	Description
read_pdb_file()	Read PDB topology file.
read_traj_file()	Read DCD MD trajectory file.
superpose_trj()	Remove roto-translational motion from MD trajectory.
encode_dcd_trajectory()	Encode trajectory with the M32K25 structural alphabet.
split_sa_align()	Split the structural alphabet alignment into regular blocks.
mi_mat()	Compute the mutual information matrix.
comp_eigensystem()	Compute the eigen system of the mutual information matrix.
block_overlap()	Compute the covariance overlap between each mutual information matrix.
matrix_smooth()	Linear smoothing of the covariance overlap blocks.
detect_sectors()	Detection of conformational sub-states.
extract_sectors()	Extract mutual information for each conformational sub-state.
allosteric_path()	Minimal distance pathfinder between allosteric and active sites.
identify_hubs()	Identify allosteric hub residues.
sectors_2dplot()	Plot the covariance overlap of time-contiguous covariance overlap.
sectors_3dplot()	Plot the covariance overlap of all combinations of time-blocked covariance overlap.

Table 5.3: Available functions contained within the R package AlloHubMat1.0

formational sub-states from multiple MD trajectories (**Fig. 5.8 A**). Extraction of correlated motions from multiple sub-states within a consistent information theoretical framework allowed us to compare the allosteric networks, both between replicas of the same liganded state and between different liganded states of PKM2. To automate the prediction of allosteric hub residues from MD trajectories, AlloMatHub was implemented as a stand-alone R package (**Fig. 5.8 B**). Here, the core functionalities of AlloHubMat will be described, followed by a discussion of how the method was used to identify allosteric residues in PKM2.

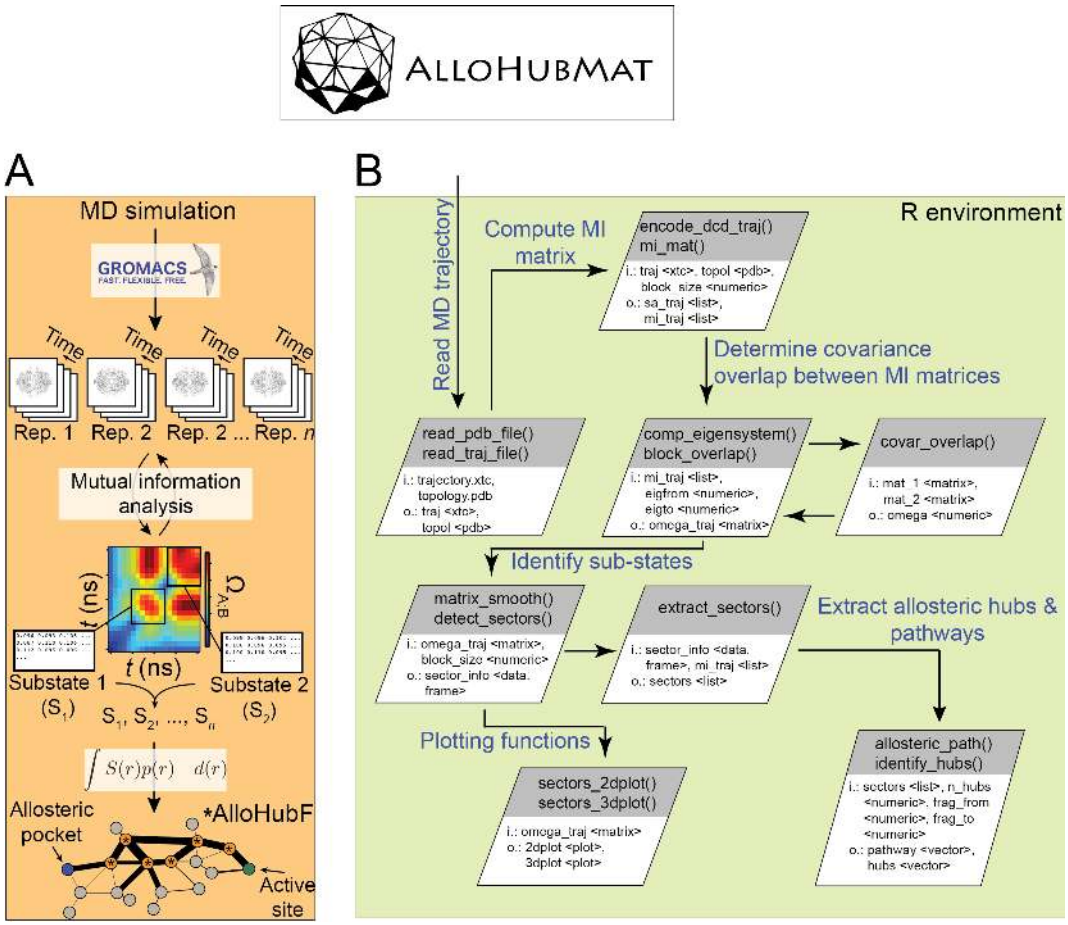

Figure 5.8: ***Schematic of the AlloHubMat and software flowchart.*** ***(A)*** *Multiple replicate molecular dynamics (MD) simulations are calculated using the GROMACS molecular dynamics engine[140]. All MD simulations are encoded with the M32K25 structural alphabet[125], and the strength of dynamic protein backbone correlations over the MD trajectory is computed using information theory mutual information statistics. The backbone correlations are explicitly used to identify and extract configurational sub-states from the MD trajectories. A global allosteric network is then constructed by integrating over the correlation matrices, and their respective probabilities, from which allosteric hub (AlloHub) fragments are extracted. AlloHub fragments are overlapping regions of four consecutive amino acid residues.* ***(B)*** *A software flowchart of AlloMatHub showing the functionalities of the software.*

5.3.1 Explicit identification of sub-states from MD simulations using information theory

Correlated motions have demonstrated as important towards protein dynamics and allostery. To extract correlated motions from MD trajectories, an information theoretical framework was used based on a coarse-grained representation of the protein backbone conformation with the M32K25 structural alphabet[125], as previously described (Section 1.3.3).

MD trajectories of tPKM2^{apo}, tPKM2^{FBP}, tPKM2$^{FBP+Ser}$ and tPKM2$^{FBP+Phe}$ were subdivided into 20 non-overlapping blocks with an equal time length of 20 ns each. For each block, the normalised mutual information between distal backbone-encoded fragments $[I^n(C_i;C_j)]$ was calculated for all pairs of fragments (i,j):

$$I_B^n(C_i;C_j) = \frac{I_B(C_i;C_j) - \epsilon_B(C_i;C_j)}{H_B(C_i,C_j)} \tag{5.10}$$

where the columns of the structural fragment alignment are given by C_i and C_j, $I(C_i;C_j)$ is the mutual information, $\epsilon(C_i;C_j)$ is the expected finite size error and $H(C_i,C_j)$ is the joint entropy (see Section 1.3.3 for a full derivation of the individual terms of the equation). The normalised mutual information $I_B^n(C_i;C_j)$ has a range between 0 and 1. Two independent, random distributions would have a normalised mutual information score of 0, whereas two identical distributions would have a normalised mutual information equal to 1. Therefore, two fragments with a normalised mutual information approaching 1 indicates that these two fragments are correlated in their motion throughout the simulation. In contrast, two fragments with a normalised mutual information approaching 0 reveals that these two fragments are uncorrelated in their motion.

With the goal of identifying conformational sub-states from a time-trajectory of mutual information matrices, eigenvalue decomposition was used to compute the geometric evolution of the protein backbone correlations. The elements of the mutual information matrix are pro-

portional to the square of the displacement, so the square root of the matrix is required to examine the extent of the matrix overlap:

$$d(A,B) = \sqrt{tr[(A^{\frac{1}{2}} - B^{\frac{1}{2}})^2]} \tag{5.11}$$

$$= \sqrt{tr[A + B - 2A^{\frac{1}{2}}B^{\frac{1}{2}}]} \tag{5.12}$$

$$= \sqrt{\sum_{i=1}^{3N}(\lambda_i^A + \lambda_i^B) - 2\sum_{i=1}^{3N}\sum_{j=1}^{3N}(\lambda_i^A \cdot \lambda_j^B)^{\frac{1}{2}}(\mathbf{v}_i^A \cdot \mathbf{v}_j^B)^2} \tag{5.13}$$

$$\Omega_{A;B} = 1 - \frac{d(A,B)}{\sqrt{trA + trB}} \tag{5.14}$$

where λ^A and λ^B denote the eigenvalues and $\mathbf{v}^A$ and $\mathbf{v}^B$ the eigenvectors of mutual information matrices A and B, N is the number of fragments used to encode the polypeptide chain. The covariance matrix overlap (Ω) ranges from 0 when matrices A and B are orthogonal, and 1 when they are identical.

The geometric difference between any two mutual information matrices, with identical dimensions, could be numerically compared by calculating the covariance overlap between the mutual information from time-contiguous trajectory blocks:

$$\Omega_{B;B-1} = 1 - \frac{d\left[I_B^n(C_i;C_j) - I_{B-1}^n(C_i;C_j)\right]}{\sqrt{tr\left[I_B^n(C_i;C_j)\right] + tr\left[I_{B-1}^n(C_i;C_j)\right]}} \tag{5.15}$$

The covariance matrix overlap between each time-contiguous mutual information matrix (Equ. 5.15) revealed an oscillatory similarity in the nMI over time (**Fig. 5.9 A** and **B**). For some regions of the MD trajectories the nMI matrices were very self-similar, reflected by $\Omega_{B;B-1} > 0.2$, suggesting that the protein was sampling a local sub-state. Conversely, other regions of the MD trajectories contained nMI matrices which were very different from the previous time block ($\Omega_{B;B-1} \to 0$), implying non-ergodic sampling. Sub-states were heuristically defined as regions which displayed Ω_i values in the top quartile (**Fig. 5.9 C** and **D**). The mutual information was averaged for each unique sub-state identified from a given trajectory, for

further analysis.

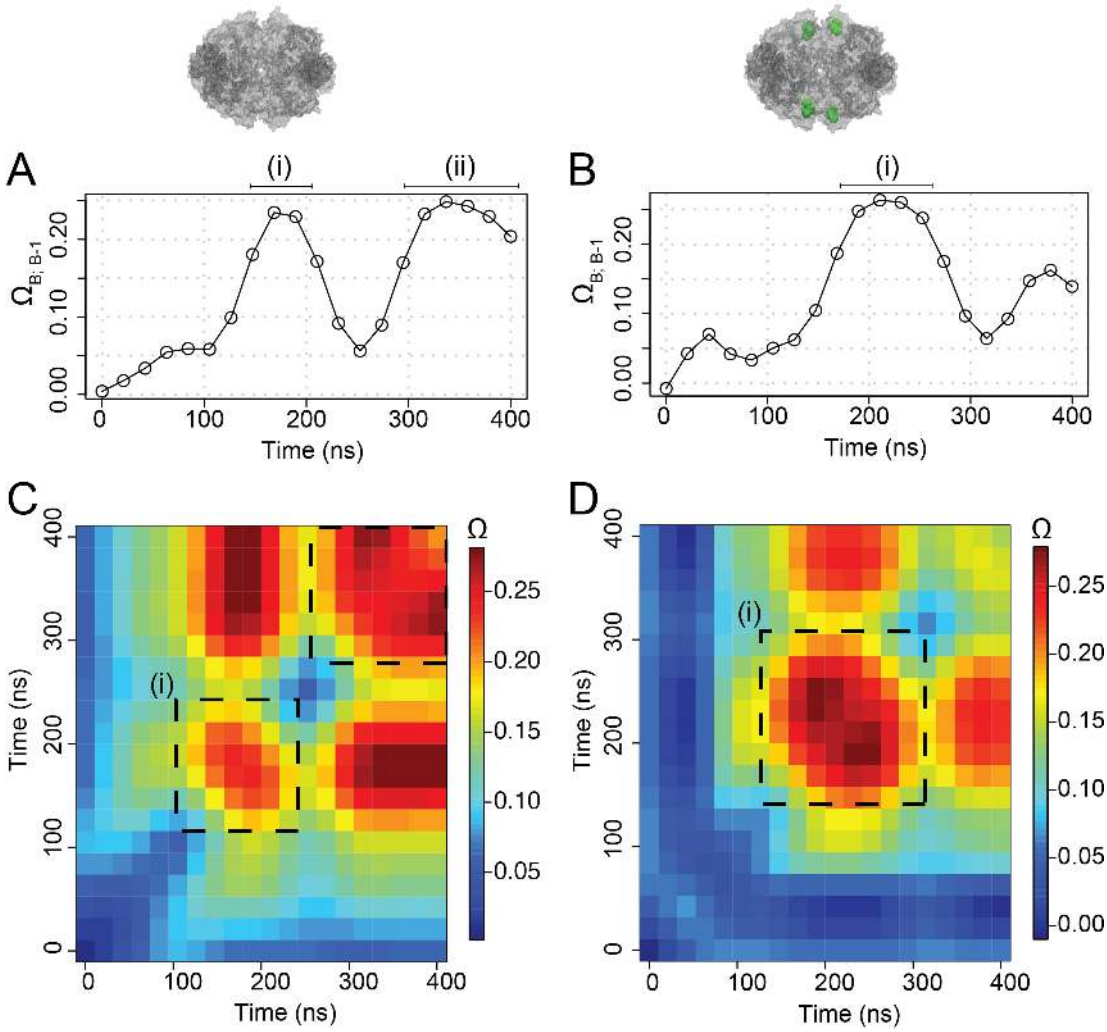

Figure 5.9: ***Conformational sub-states are identified from the geometric progression of the mutual information matrix.*** *The time-dependent matrix covariance overlap between contiguous mutual information matrices computed from contiguous trajectory blocks is quantified using Equ. 5.15 for an example replica of* ***(A)*** *$tPKM2^{Apo}$ and* ***(B)*** *$tPKM2^{FBP}$. The matrix covariance overlap was computed for all combinations of time-resolved nMI blocks. The resulting matrices were smoothed and are shown for example replicate MD simulations of* ***(C)*** *$tPKM2^{apo}$ and* ***(D)*** *$tPKM2^{FBP}$. Sub-states were identified as regions containing a covariance overlap in the top quartile of the trajectory (shown as dashed boxes).*

5.3.2 PKM2 sub-states cluster according to the liganded state of the MD simulation

A total of 7 substates were identified for all MD simulations of tPKM2^{apo}, 6 for tPKM2^{FBP}, 4 for tPKM2$^{FBP+Ser}$ and 3 for tPKM2$^{FBP+Phe}$. To investigate whether the correlated motions for each sub-state could be attributed to the liganded state of the MD simulation, the mutual information matrices extracted from each unique sub-state were compared with a complete-linkage hierarchical clustering, using the covariance matrix overlap in Equ. 5.14 as a distance metric (**Fig. 5.10**). Clustering of the sub-states revealed four predominant clusters (denoted as *C1 - C4*). nMI matrices from tPKM2^{apo} were found to predominate in cluster 1, tPKM2^{FBP} in cluster 2 and tPKM2$^{FBP+Ser}$ in cluster 3. We found cluster 4 to be populated by both tPKM2^{apo} and tPKM2$^{FBP+Phe}$, as well as a number of tPKM2^{FBP} replicas.

The commonality between the sub-state mutual information matrices of the same liganded state suggested that each of the replicate MD simulations converged to a common sub-state, which was ligand-dependent. Moreover, this suggested that allosteric ligand-dependent correlated motions were captured by the preceding analysis. Therefore, an ensemble-averaged nMI matrix was computed for PKM2 in each of the four liganded states, as an integral over all sub-states (I^n_{ss}) of the protein (r) weighted by the probability of the sub-state (p):

$$\langle I^n_{ens} \rangle = \int I^n_{ss}\, p(r) \quad dr \tag{5.16}$$

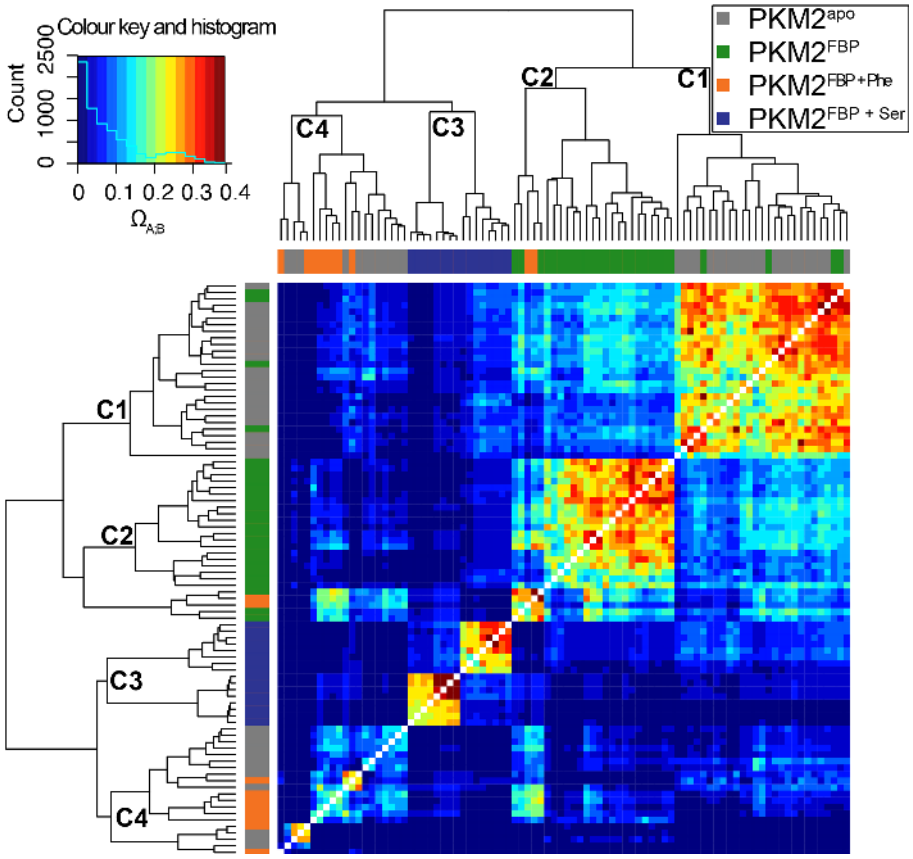

Figure 5.10: ***PKM2 correlated motions cluster according to the liganded state of the simulation.*** *A complete-linkage hierarchical clustering of mutual information matrices, computed from simulated trajectories of $tPKM2^{apo}$ (grey), $tPKM2^{FBP}$ (green), $tPKM2^{FBP+Ser}$ (blue) and $tPKM2^{FBP+Phe}$ (orange). Four clusters are assigned C1 - C4. For every sub-state, the network of correlations from each of the four protomers is presented individually.*

5.3.3 A disperse network of hub residues are predicted to propagate FBP-induced activation of PKM2

The ensemble-averaged matrix of correlated motions identified from the analysis of tPKM2^{apo} were subtracted from those of tPKM2^{FBP} to isolate the fragment positions predicted to be involved in the allosteric state transition, which were termed allosteric hub fragments (Allo-HubFrags). An inspection of the frequency distribution of the mutual information matrices of tPKM2^{apo} and tPKM2^{FBP} found that the strength of the fragment correlations were log-normally distributed with a large Gaussian component, and a long tail to the distribution (**Fig. 5.11 A** and **B**). Subtracting the mutual information matrices resulted in the z-transformation of the distribution of correlations, with a mean $\mu_{I^n} = 0$ and a standard deviation $\sigma_{I^n} = 0.04$ (**Fig. 5.11 C**).

The variance of each of the 2119740 individual backbone couplings were computed within the multiple conformational sub-states for tPKM2^{apo} and tPKM2^{FBP}. We found that when the average difference in the mutual information between the two liganded states ($\mu\Delta^{FBP}$) was inspected against the variance of each of the couplings ($\sigma\Delta^{FBP}$), the resulting plot formed a v-shape with highly correlated fragments showing a high degree of variance (**Fig. 5.11 D**).

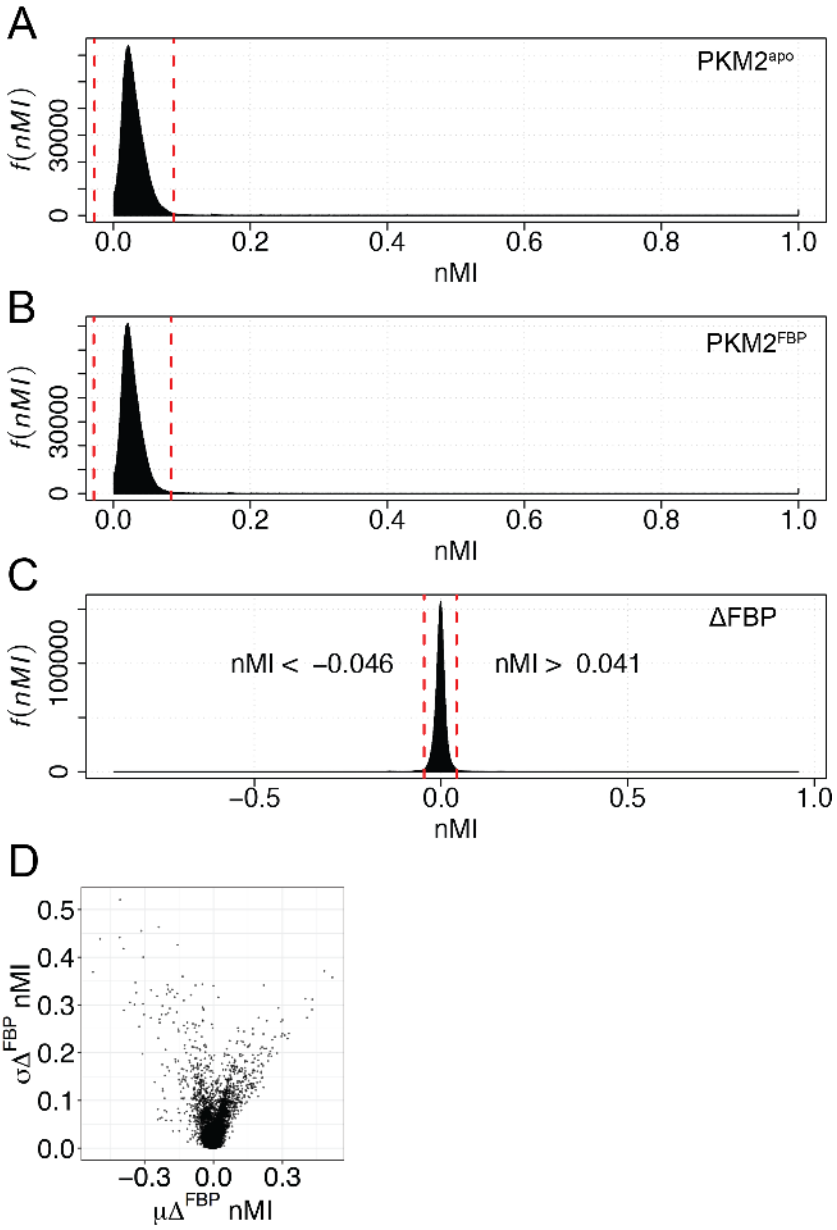

Figure 5.11: ***The sub-state mutual information matrix is log-normally distributed.*** *The distribution of mutual information couplings for the ensemble-averaged mutual information matrices of* ***(A)*** $tPKM2^{apo}$, ***(B)*** $tPKM2^{FBP}$, *and* ***(C)*** *the subtracted mutual information matrices of the preceding two liganded states.* ***(D)*** *The variance of each coupling was computed over each of the multiple conformational sub-states, for each fragment coupling.*

Rather than ranking fragment positions by their nMI score, the significance of the change in nMI was computed for each coupling (**Fig. 5.12 A**). From a total of 76 fragments identified as significant, the top AlloHubFrags were selected for further analysis ($Hubs_{1-10}$). Hub_5 and Hub_6 were found to overlap with a two amino acid residue difference between the two hubs. The AlloHubFrags were found to be spatially diverse across the PKM2 structure, though not obviously structurally contiguous (**Fig. 5.12 B**). Hub_5 and Hub_6 were proximal to the A-A' interface, Hub_9 was proximal to the FBP binding pocket, Hub_{10} localised to the C-C' interfaces and Hub_1 and Hub_2 were in the B-domain (**Fig. 5.12 B**).

The subtracted mutual information matrix was next distance-weighted using the Cα distance matrix (M) determined from the PKM2 crystal structure:

$$I^n(C_i; C_j)' = I^n(C_i; C_j) \cdot \frac{1}{M} \tag{5.17}$$

Re-weighting the mutual information matrix with the inverse of the distance matrix had the effect of down-weighting long-distance correlations and up-weighting short-range correlation, along which non-bonded chemical interactions could occur. From the distance-weight mutual information matrix, minimal distance pathways between FBP binding-pocket fragments and active-site fragments were computed using the Dijkstra algorithm in order to predict the allosteric pathways elicited by FBP. This analysis revealed that all AlloHubFrags, with the exception of Hub_5 and Hub_6, were connected to the predicted pathways (**Fig. 5.12 B**), supporting the hypothesis that the AlloHubFrags propagate the allosteric effect of FBP.

Of the 40 residues within the top ten-ranked AlloHubFrags, we sought to rationally guide the selection of a number of residues for experimental mutagenesis. Each fragment used to encode the backbone conformation of the MD trajectories of PKM2 is a collection of four contiguous residues, and as such, the mutual information between two fragments could not be readily assigned to an individual amino acid residue. Nevertheless, the chain rule for mutual

information states that the mutual information of a collection of random variables is the sum of the conditional mutual informations:

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, X_{i-2}, ..., X_1) \tag{5.18}$$

Assuming that the variables $X_1, X_2, ..., X_n$ are random, it would be possible to reconstruct the per-residue mutual information and thus select top-ranked residues with high mutual information content for mutagenesis. The validity of this assumption was tested by computing the distribution of mutual information between neighbouring fragments (**Fig. 5.12 C**). We observed a distribution of mutual information between neighbouring fragments, which was skewed towards higher values, when compared to the same log-normal distribution of the entire mutual information matrix, suggesting that the majority of neighbouring fragments were correlated and were non-random. This proved that the underlying assumption of Equ. 5.18 did not hold. Thus, it was not possible to numerically assign a per-residue mutual information score.

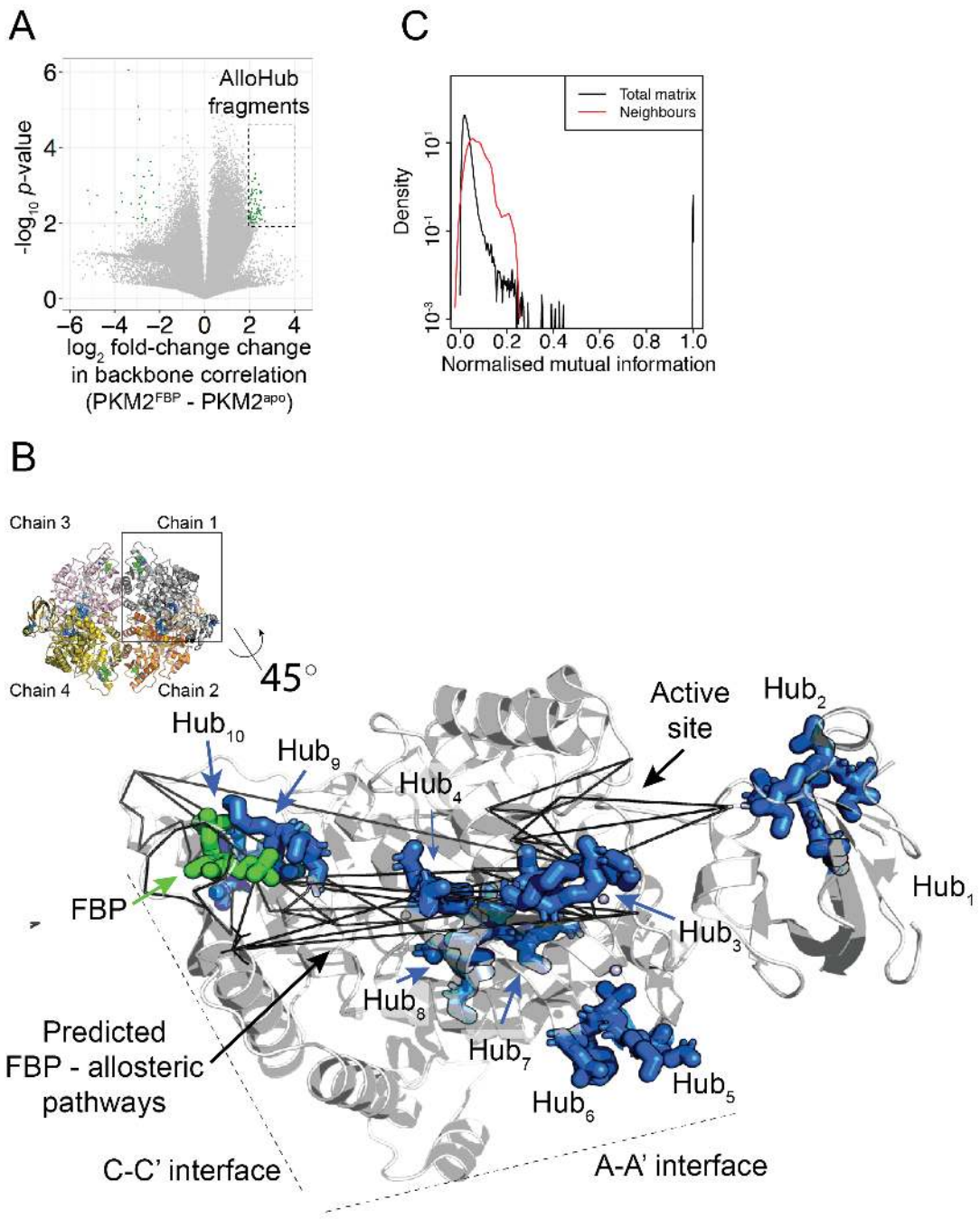

Figure 5.12: ***Hub residues are predicted to propagate energy transfer along allosteric pathways.*** ***(A)*** *A volcano plot showing the difference in mutual information couplings between $tPKM2^{apo}$ and $tPKM2^{FBP}$.* ***(B)*** *The spatial distribution of the top ten predicted AlloHubFrags projected onto a cartoon representation of PKM2. Predicted allosteric pathways between the FBP binding pocket and the active site are shown as black lines.* ***(C)*** *The distribution of normalised mutual information for all fragment combinations (black) and for neighbouring fragments (red).*

5.3.4 Design of allosteric hub mutants (AlloHubMs)

In lieu of an analytic solution to the per-residue mutual information, allosteric hub mutants (AlloHubMuts) were generated by substituting AlloHubFrag residues with amino acids that were predicted to be tolerated at the respective position based on their occurance in a multiple sequence alignment of 5381 pyruvate kinase homologues. Of the residues within the top-ten AlloHubFrags, there was variability in the degree of sequence conservation. Hub_7 and Hub_8 contained several highly conserved hydrophobic residues on parallel β-sheets within the A-domain of the protein. Of these V324, I325, A327, G355 and D357 were found to be very highly conserved within the alignment (**Fig. 5.13 A**). Conversely, residues R489 - F492 within Hub_{10}; and residues T121 - I124 within Hub_1 and L203 - K206 Hub_2 in the B-domain, were highly sequence variable (**Fig. 5.13 A**) and are exposed in the crystal structure of human PKM2 (**Fig. 5.13 B**).

Amount a total of 32 PKM2 AlloHubMuts generated, we chose to experimentally characterise seven variants [I124G, F244V, K305Q, F307P, A327S, C358A, R489L (**Fig. 5.13 B**)] because they expressed as soluble protein and had very similar secondary structure profiles to that of PKM2(WT).

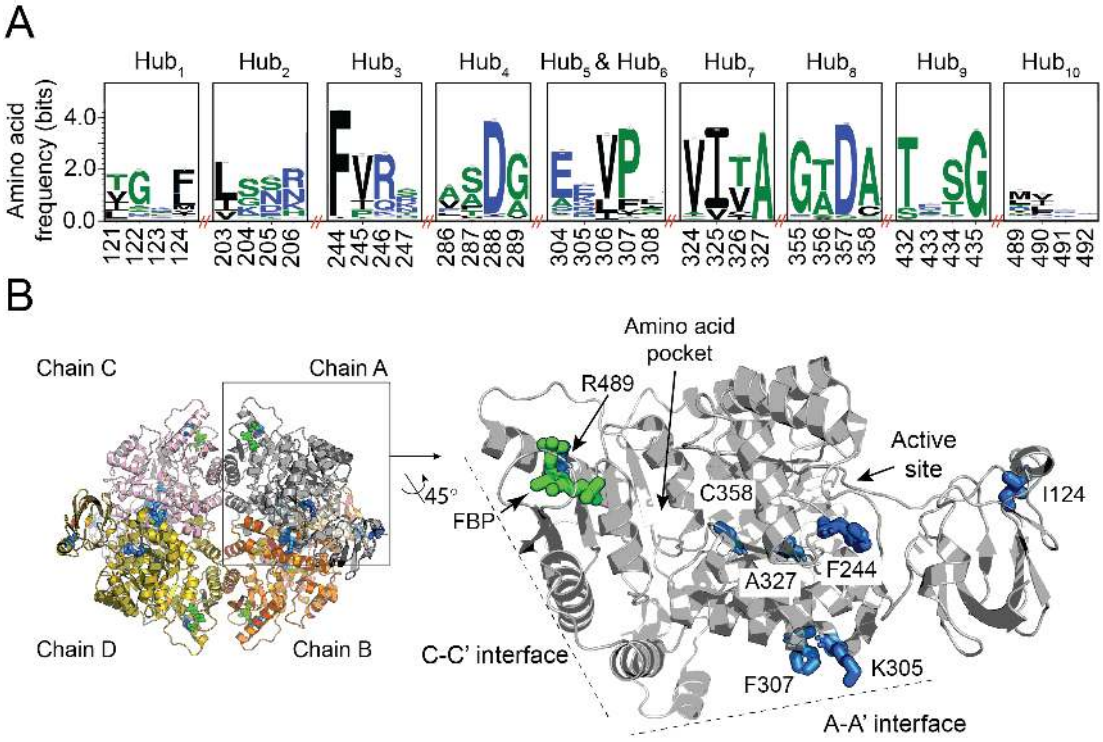

Figure 5.13: ***Design of Allosteric Hub Mutants (AlloHubMuts).*** *Allosteric Hub Mutants (AlloHubMuts) were designed based on an empirical prediction of which residues would be tolerated at each AlloHubFrag position.* ***(A)*** *A logo-plot of the AlloHubFrags from a multiple sequence alignment of 5381 pyruvate kinase homologues.* ***(B)*** *The residue positions of the AlloHubMuts.*

5.4 Conclusion

Ion-mobility coupled to mass spectrometry measurements in Chapter 4 revealed that PKM2 undergoes conformational changes upon FBP binding, though the molecular details of the conformational changes were elusive. To this end, we performed an extensive characterisation of the conformational dynamics of PKM2 bound to various allosteric ligands. MD simulations of monomeric and tetrameric PKM2 found that allosteric activator binding induces the closure of the B-domain over the catalytic pocket. The ligand-dependent dynamics of the B-domain is therefore proposed as a determinant of enzyme activation, which traps highly resident water molecules proximal to the substrate binding pocket.

A computational method *AlloHubMat* was developed and applied towards identifying allosteric hub residues, which orchestrate the transmission of allosteric information between ligand binding pockets and the active site. We found a network of allosteric hub residues, which are predicted to potentiate the effects of FBP allosteric within PKM2. The AlloHubFrags were used as a template to design a collection of single-point mutant variants (AlloHubMuts) of PKM2, to test the hypothesis that these residues are involved in the allosteric mechanism of the protein.

Chapter 6

Allosteric mutants reveal distinct ligand pathways that combinatorially regulate PKM2.

6.1 Introduction

Chapter 5 described a novel computational method *AlloHubMat* to predict allosteric hub residues from an ensemble of conformational sub-states obtained from multiple replicate molecular dynamics (MD) simulations. The method was applied towards analysing MD trajectories of tPKM2^{apo} and tPKM2^{FBP}, revealing a network of residues that are predicted to propagate FBP allostery. We hypothesised that changing the side-chain chemistry of a selection of these AlloHub residues would perturb allosteric coupling between the FBP pocket and the active site, thus validating the involvement of these residues in the allosteric mechanism of PKM2.

To this end we designed a selection of AlloHub mutants (AlloHubMuts; I124G, F244V, K305Q, F307P, A327S, C358A, R489L) and sought to comprehensively characterise their biophysical and enzymatic behaviour using an integrative experimental strategy detailed herein (**Fig. 6.1**).

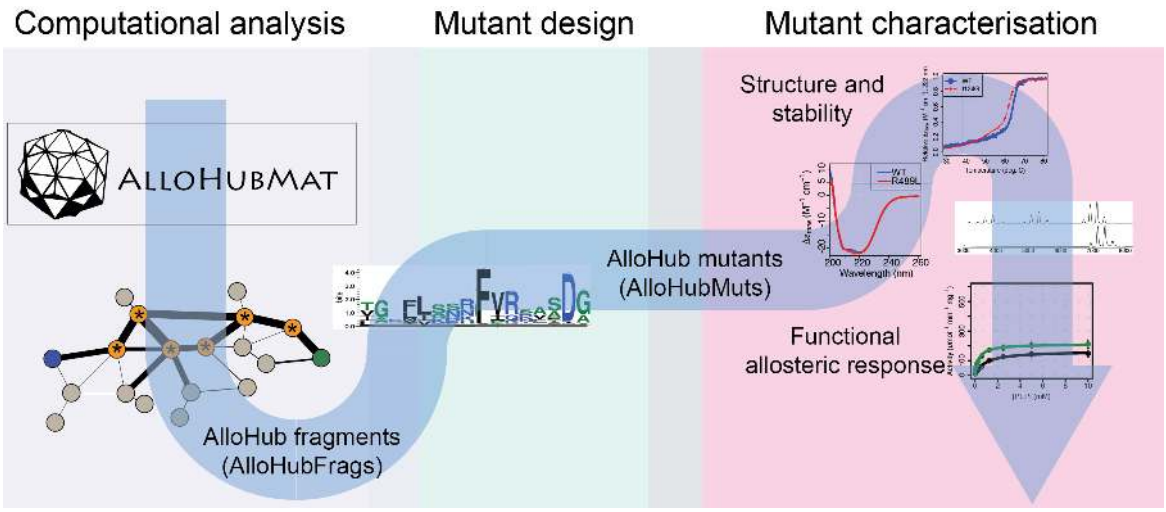

Figure 6.1: ***An integrated computational and experimental strategy for identifying residues which transmit allosteric information between a binding pocket and the active site of a protein.*** *AlloHubMat is used to identify candidate allosteric residues, which guide the rational design of allosteric hub mutants (AlloHubMuts). The AlloHubMuts are characterised with a series of experimental methods, in order to interrogate their biophysical and enzymatic response to FBP, Phe and Ser.*

6.2 Biophysical characterisation of the AlloHubMuts

6.2.1 AlloHubMuts have the same secondary structure content as PKM2(WT)

Small changes to the local chemical composition of proteins can result in significant conformational changes. Even single-point mutants can affect the stability of proteins, resulting in changes to secondary structural content[178]. In order to investigate the secondary structure content of the AlloHubMuts and to determine whether the inserted amino acid changes significantly affected the protein fold, far-UV circular dichroism (CD) spectra were acquired for PKM2(WT) and for each of the AlloHubMuts. A superimposition of the AlloHubMut far-UV CD spectra with that of PKM2(WT) revealed very little detectable difference in the shape and intensities of the spectra (**Fig. 6.2**), suggesting that none of the AlloHubMuts appreciably changed the secondary structure content of PKM2.

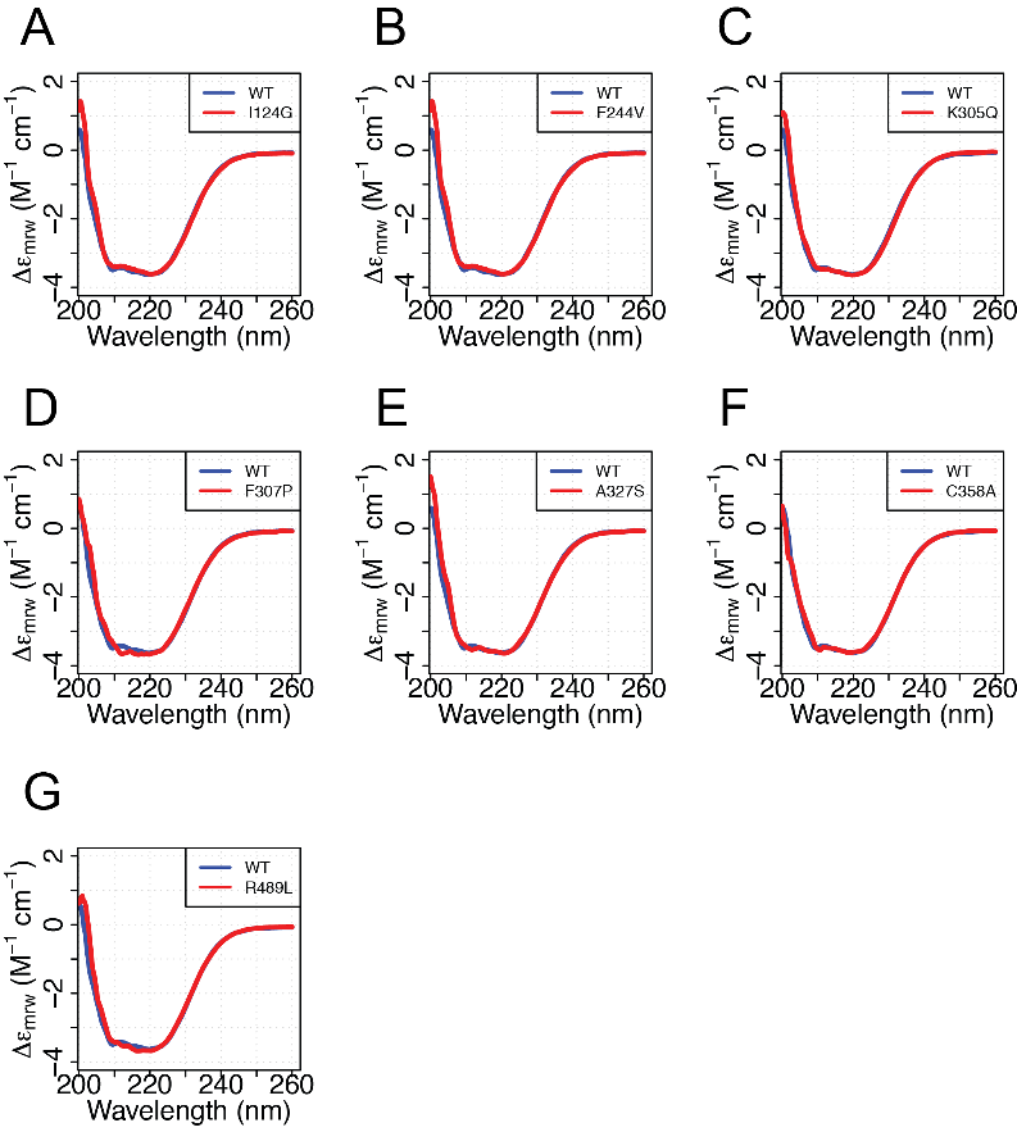

Figure 6.2: ***Far-UV circular dichroism spectra of AlloHubMuts suggest no change in secondary structure content.*** *Far-UV CD spectra of PKM2(WT) and of the seven AlloHubMuts were acquired between 200 nm and 260 nm. Spectra of the AlloHubMuts (red) are superimposed with that of PKM2(WT) (blue);* ***(A)*** *I124G,* ***(B)*** *F244V,* ***(C)*** *K305Q,* ***(D)*** *F307P,* ***(E)*** *A327S,* ***(F)*** *C358A,* ***(G)*** *R489L. A protein concentration of 0.15 mg · mL^{-1} was used in the acquisition of all spectra.*

6.2.2 AlloHubMuts at either the A-A' or C-C' interfaces show distinct thermodynamics properties

The far-UV CD spectra of the AlloHubMuts were similar to that of PKM2(WT), suggesting that the mutants did not significantly alter the secondary structure of the protein. Nevertheless, a substantial alteration in secondary structure content would likely be required for changes to become apparent in the far-UV CD spectrum, given the large size of the PKM2 tetramer (234 kDa). Moreover, local conformational changes could conceivably perturb protein stability without enacting changes to the secondary structure content.

To this end, thermal unfolding experiments were performed to determine the melting temperature (T_m) of each AlloHubMut. PKM2(K305Q) and PKM2(R489L) were found to be thermally destabilising, reflected by a significant decrease in the T_m relative to PKM2(WT) (**Fig. 6.3** and Table 6.1). None of the other AlloHubMuts were found to have a significant effect on the melting temperature, compared to PKM2(WT) (**Fig. 6.3 A-G**). The two destabilising mutants K305Q and R489L are located at subunit interfaces within the PKM2 structure. K305 forms a charged interaction with E384 on the protomer on the opposite side of the A-A' interface (**Fig. 6.3 H**). Replacing the ϵ-ammonium group of lysine with an amide group of glutamine caps the side-chain charge and likely prevented the formation of inter-protomeric charged interactions at position 305, possibly reducing oligomer formation and stability of the protein. R489 is positioned proximal to the C-C' interface, though does not form inter-protomeric interactions in the crystal structure of PKM2. Rather, its guanidino group forms charged interactions with the 1'-phosphate group of FBP (**Fig. 6.3 I**), and therefore contributes to the binding of FBP. Replacing the guanidino group with an isobutyl group at position 489 likely reducing the binding affinity of FBP.

Ligand binding can increase a protein's thermal stability because the ligand favours the folded state in the folded-to-unfolded equilibrium. Nevertheless, we consistently found that the addition of saturating concentrations of FBP to PKM2(WT) did not change the melting temper-

ature of the protein (**Fig. 6.4 A**). The lack of FBP-induced thermal stabilisation was likely due to co-purification of PKM2(WT) with endogenous FBP (Section 3.2.1). Similarly, I124G, F244V, F307P, A327S and C358A showed no significant changes in their thermal stabilities upon FBP addition (**Fig. 6.4** and Table 6.1), suggesting that these variants also retain sub-stoichiometric amounts of endogenous FBP during protein purification.

The mutation with the largest destabilising effect was PKM2(K305Q), which produced a 19.1 °C decrease in the apparent melting temperature and a 174.9 (kJ/mol) decrease in the change in enthalpy at T_m, compared to the wild-type. In contrast, the PKM2(R489L) at the C-C' interface produced a comparatively small decrease in the melting temperature (5.6 °C). Intriguingly, the thermal stability of the R489L mutant was rescued the $\Delta T_m^{WT \rightarrow R489L}$ from 5.6 to 2.4 °C - equivalent to wild-type levels (**Fig. 6.4** and Table 6.1). In contrast, even when in the presence of 1 mM FBP, PKM2(K305Q) was markedly less stable than the wild-type protein (**Fig. 6.4** and Table 6.1).

Taken together, experimental analyses of the thermal stability of the AlloHubMuts revealed that all AlloHubMuts, with the exception of K305Q and R489L, had similar equilibrium unfolding kinetics as PKM2(WT). Given that the position of R489 is proximal to the FBP binding pocket and the finding that FBP addition rescued the thermal stability of the mutant variant, it is likely that PKM2(R489L) reflects the thermal stability of PKM2 devoid of FBP. Conversely, PKM2(K305Q) is hypothesised to introduce changes to the A-A' interface charged interactions, resulting in protein destabilisation, which is not rescued by FBP addition.

PKM2 variant	ΔH (kJ/mol)	ΔS (kJ/K · mol)	T_m (°C)
WT	525.8 ± 44.8	8.2 ± 0.7	63.9 ± 0.1
WT + 1 mM FBP	492.4 ± 23.6	7.8 ± 0.4	63.1 ± 0.1
I124G	336.5 ± 23. 1	5.4 ± 0.4	61.3 ± 0.1
I124G + 1 mM FBP	600.0 ± 22.6	9.8 ± 0.4	61.3 ± 0.4
F244V	477.2 ± 16.9	7.6 ± 0.3	62.5 ± 0.1
F244V + 1 mM FBP	525.0 ± 34.5	8.6 ± 0.6	61.3 ± 0.1
K305Q	350.9 ± 13.5	7.8 ± 0.3	44.8 ± 0.1
K305Q + 1 mM FBP	479.5 ± 20.9	8.7 ± 0.4	54.8 ± 0.1
F307P	477.8 ± 18.2	7.7 ± 0.3	62.3 ± 0.1
F307P + 1 mM FBP	695.2 ± 36.2	11.3 ± 0.6	61.7 ± 0.1
A327S	558.2 ± 14.8	9.2 ± 0.2	60.8 ± 0.4
A327S + 1 mM FBP	653.4 ± 40.8	10.9 ± 0.7	59.9 ± 0.1
C358A	493.9 ± 20.1	7.9 ± 0.3	62.7 ± 0.1
C358A + 1 mM FBP	660.4 ± 22.2	11.0 ± 0.4	60.3 ± 0.1
K433E	488.4 ± 9.6	5.1 ± 0.2	56.8 ± 0.1
K433E + 1 mM FBP	678.0 ± 27.6	11.6 ± 0.5	58.1 ± 0.1
R489L	232.9 ± 21.4	4.0 ± 0.4	58.2 ± 0.3
R489L + 1 mM FBP	643.3 ± 45.6	10.5 ± 0.7	61.5 ± 0.1

Table 6.1: ***Equilibrium thermal unfolding kinetic parameters of the AlloHubMuts.*** *A two-phase unfolding model (Methods Section 2.3.6) was fit to each of the unfolding curves shown in* ***Fig. 6.3****. The change in enthalpy (ΔH) and entropy (ΔS) between the folded and semi-unfolded states, as well as the melting temperature (T_m) is reported.*

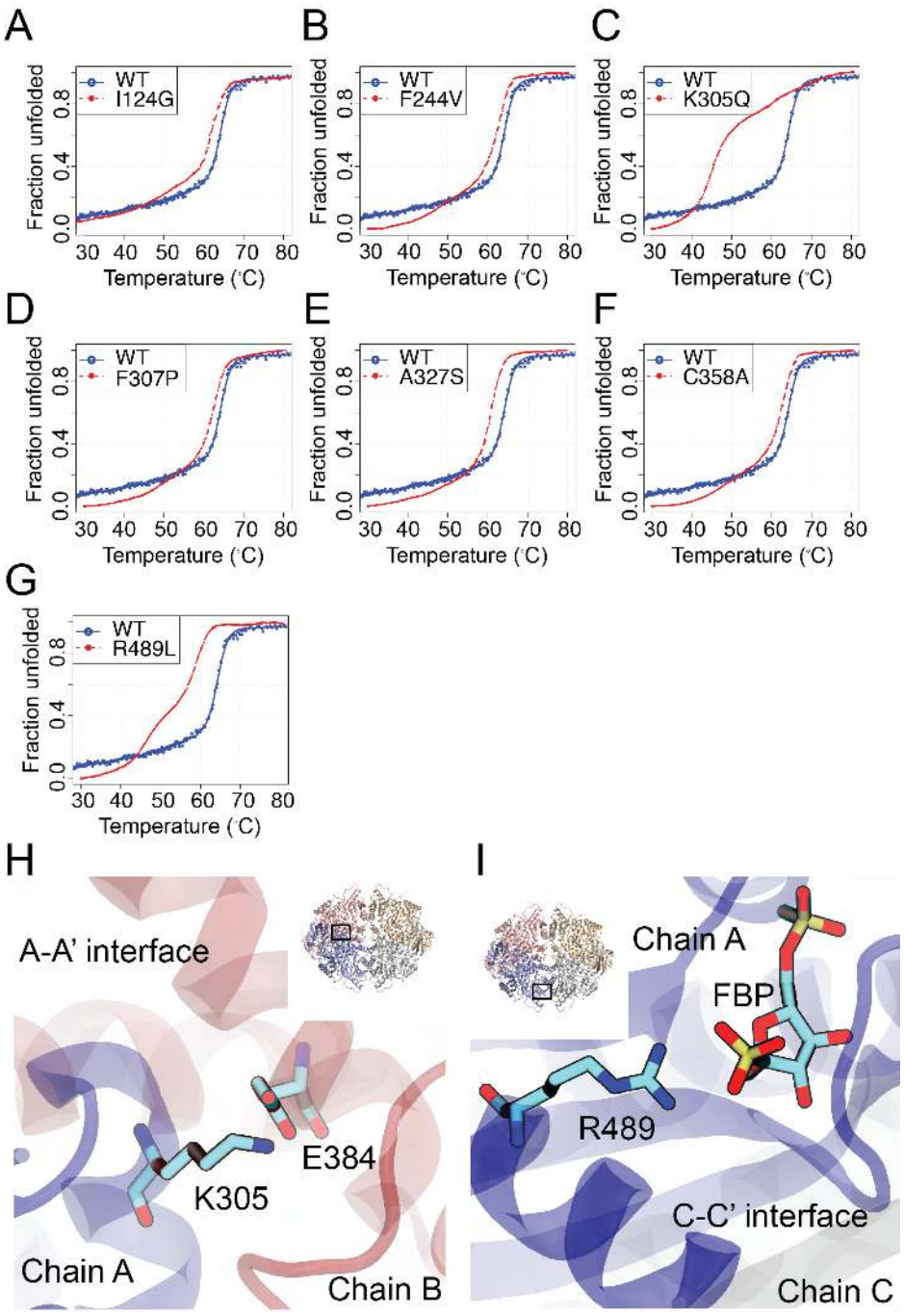

Figure 6.3: ***Thermal unfolding spectra of the AlloHubMuts fitted to a two-phase unfolding model.*** ***(A-G)*** *Far-UV CD intensity at 222 nm of PKM2(WT) (blue) and the AlloHubMuts (red) were monitored over a range of temperatures between 30 °C and 80 °C, in the absence of any added ligands. Data points were fitted to a two-phase unfolding curve.* ***(H)*** *A schematic of the local structure of K305 at the A-A' interface and* ***(I)*** *R489 at the FBP binding pocket.*

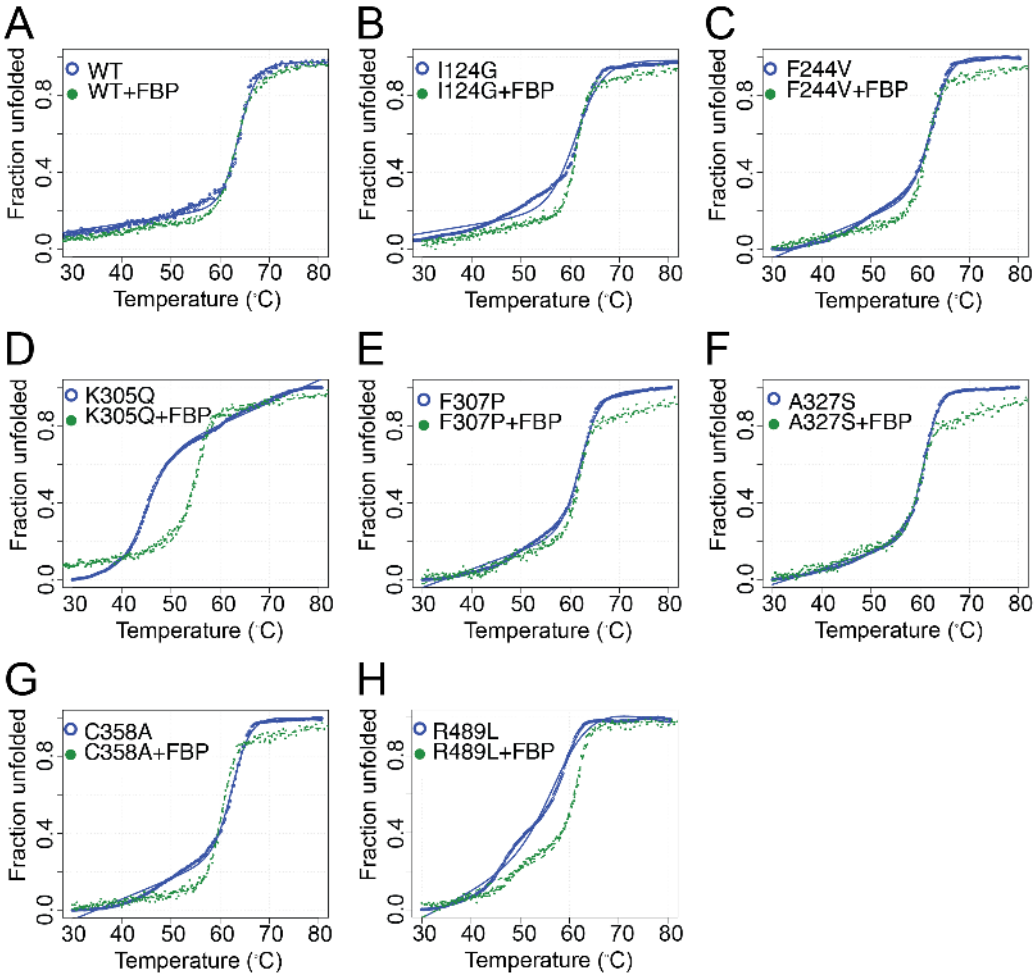

Figure 6.4: ***The thermal stability of PKM2(K305Q) and PKM2(R489L) is increased upon FBP addition.*** *Far-UV CD intensity at 222 nm was monitored for* ***(A)*** *PKM2(WT) and* ***(B-H)*** *the AlloHubMuts in the absence (blue) and in the presence (green) of 1 mM FBP.*

6.2.3 R489L reduces the binding affinity to FBP

The reduced thermal stability of PKM2(R489L) and the fact that R489 forms a charge-charge interaction with the 1'-phosphate group of FBP, suggested that this mutant may have a reduced FBP binding affinity. Binding affinities of the AlloHubMuts to FBP (Table 6.2) were calculated from fluorescence titration measurements (**Fig. 6.5 A-G**), as previously described (Section 3.2.1). All of the AlloHubMuts bound to FBP with nano molar affinity, with the exception of PKM2(R489L), which was estimated to bind to FBP with an apparent affinity of (14.0 ± 2.7) mM. Despite the two-order of magnitude decrease in the apparent binding affinity of PKM2(R489L) for FBP, ligand binding saturated the measured tryptophan fluorescence changes at mM concentrations (**Fig. 6.5 G**).

PKM2	K_D^{FBP}
WT	(21.4 ± 9.0) nM
I124G	(39.5 ± 33.5) nM
F244V	(30.7 ± 33.1) nM
K305Q	(39.4 ± 34.1) nM
F307P	(4.0 ± 12.8) nM
A327S	(43.2 ± 48.7) nM
C358A	(35.3 ± 23.6) nM
R489L	(14.0 ± 2.7) mM

Table 6.2: ***Steady-state dissociation constants of the AlloHubMuts for FBP.*** *A 1:1 stoichiometry binding curve (3.2.1) was fit to the binding curves in* ***Fig. 6.5****. The K_D^{FBP} is reported in variable units.*

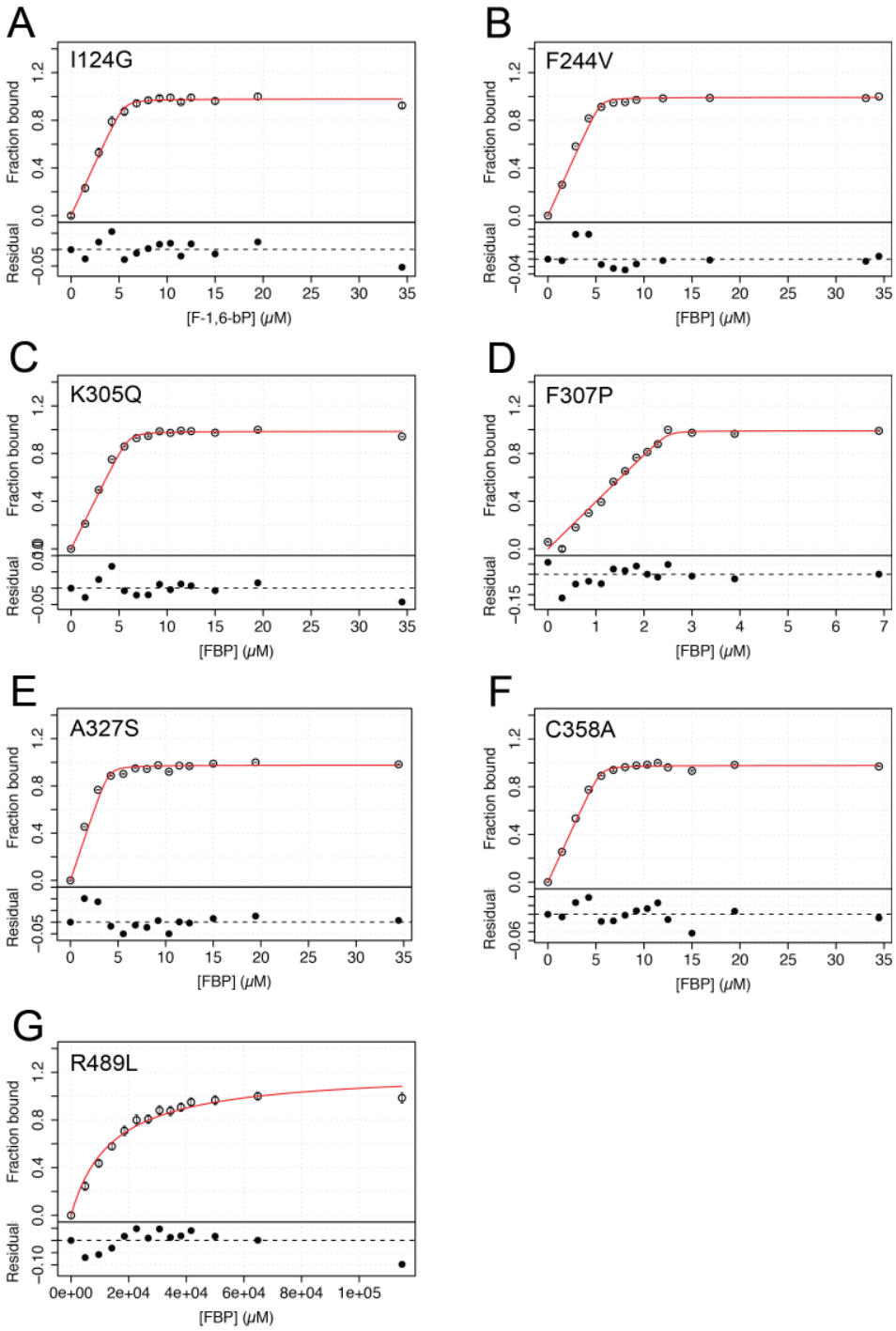

Figure 6.5: ***The AlloHubMuts have a nano molar affinity for FBP with the exception of PKM2(R489L).*** *The ratio of the fluorescence emission intensity at 325 nm and 350 nm were recorded over a range of FBP concentrations. Apparent binding curves were fit assuming a 1:1 stoichiometry (red).*

6.3 Characterisation of the functional response of AlloHubMuts to allosteric ligands

The AlloHubMuts were identified by the software AlloHubMat as candidate residues, which coordinated the allosteric effect of FBP. It was postulated that changing the chemical composition of these residues would perturb the allosteric coupling between the FBP binding pocket and the active site. To test this hypothesis we sought to characterise the steady-state enzyme kinetics of each AlloHubMut and the activity response of each mutant to the addition of FBP.

6.3.1 Several AlloHubMuts attenuate the allosteric coupling between the FBP pocket and the active site

To assess the functional effect of perturbing the *in silico*-determined allosteric pathways, we measured the steady-state kinetics of the AlloHubMuts in the absence and in the presence of saturating concentrations of FBP. Initial velocity curves were determined over a range of concentrations of phosphoenolpyruvate and revealed that all of the AlloHubMuts had a lower apparent maximal velocity, compared to PKM2(WT), in the absence of FBP (Table 6.3). The addition of FBP resulted in a marked increase in the apparent maximal velocity of PKM2(I124G) and PKM2(R489L) (Table 6.3). PKM2(K305Q) was inactive, both in the absence and in the presence of FBP, suggesting that this AlloHubMut was catalytically dead (Table 6.3). In contrast, PKM2(F307P) appeared to be constitutively active, with a low apparent K_M^{PEP}, which was unchanged upon FBP addition (Table 6.3).

As described previously in Chapter 3, FBP was found to exert its functional allosteric effect on PKM2(WT) by increasing the enzyme-substrate affinity, without changing the maximal velocity. It was therefore postulated that allosteric activation results from a coupling between FBP and PEP binding, mediated by the AlloHub residues (**Fig. 6.6 A**). Consequently, we sought to quantitatively determine the coupling between the FBP pocket and the active site for PKM2(WT) and each of the AlloHubMuts, to discern whether, and to which extent, mutations along the predicted FBP pathway perturbed the allosteric coupling.

Assuming a single-substrate-single-modifier mechanism of PKM2 catalysis (**Fig. 6.6 B**), the coupling between the allosteric pocket and the substrate binding site was determined by calculating the log-fold ratio of the Michaelis-Menten constant of the protein saturated with FBP (K_M^0), divided by Michaelis-Menten constant in the absence of FBP (K_M^X)[152]:

$$Q = \left(\frac{\alpha K_S}{K_S} \right) \tag{6.1}$$

A $log\, Q > 0$ indicates a positive coupling (activation) and $log\, Q < 0$ negative coupling (inhibition). AlloHubMuts I124G, F244V, K305Q, F307P and R489L were all found to show attenuated allosteric coupling between the active site and the FBP pocket compared to PKM2(WT) (**Fig. 6.6 C**), indicating that *AlloHubMat* identified residues that mediate the allosteric effect of FBP. In contrast, AlloHubMuts A327S and C358A both had a coupling co-efficient (Q), which was unchanged from that of WT, suggesting either that the amino acid substitutions were functionally neutral, or that these residues are not required for FBP-induced activation.

6.3.2 AlloHub residues A327 and C358 mediate multi-ligand allosteric coupling

Previous work presented in Chapters 3 and 4, found that the allosteric inhibitor phenylalanine (Phe) acts to reduce the activating effect of FBP through a functional cross-talk. We sought to investigate whether AlloHubMuts, which abrogated FBP-induced allostery, also perturbed the cross-talk between FBP and Phe.

To this end, the coupling co-efficient was measured for the doubly liganded $PKM2^{FBP+Phe}$ state ($Q^{FBP+Phe}$). AlloHubMuts I124G, F244V and R489L revealed the same $Q^{FBP+Phe}$ as the wild-type variant (**Fig. 6.6 D**), suggesting either that the allosteric coupling between FBP, Phe and the active site was unchanged for these mutant substitutions, or that residues I124, F244 and R489 do not contribute to the FBP-Phe cross-talk.

In contrast, the simultaneous addition of FBP and Phe failed to produce an allosteric response for AlloHubMuts K305Q and F307P (**Fig. 6.6 D**), further suggesting that these variants are allosterically inert. Moreover, the addition of Phe to A327S and C358, failed to attenuate FBP-induced activation of these AlloHubMuts, indicating that residues A327 and C358 have a role in coupling the allosteric effect of Phe with that of FBP.

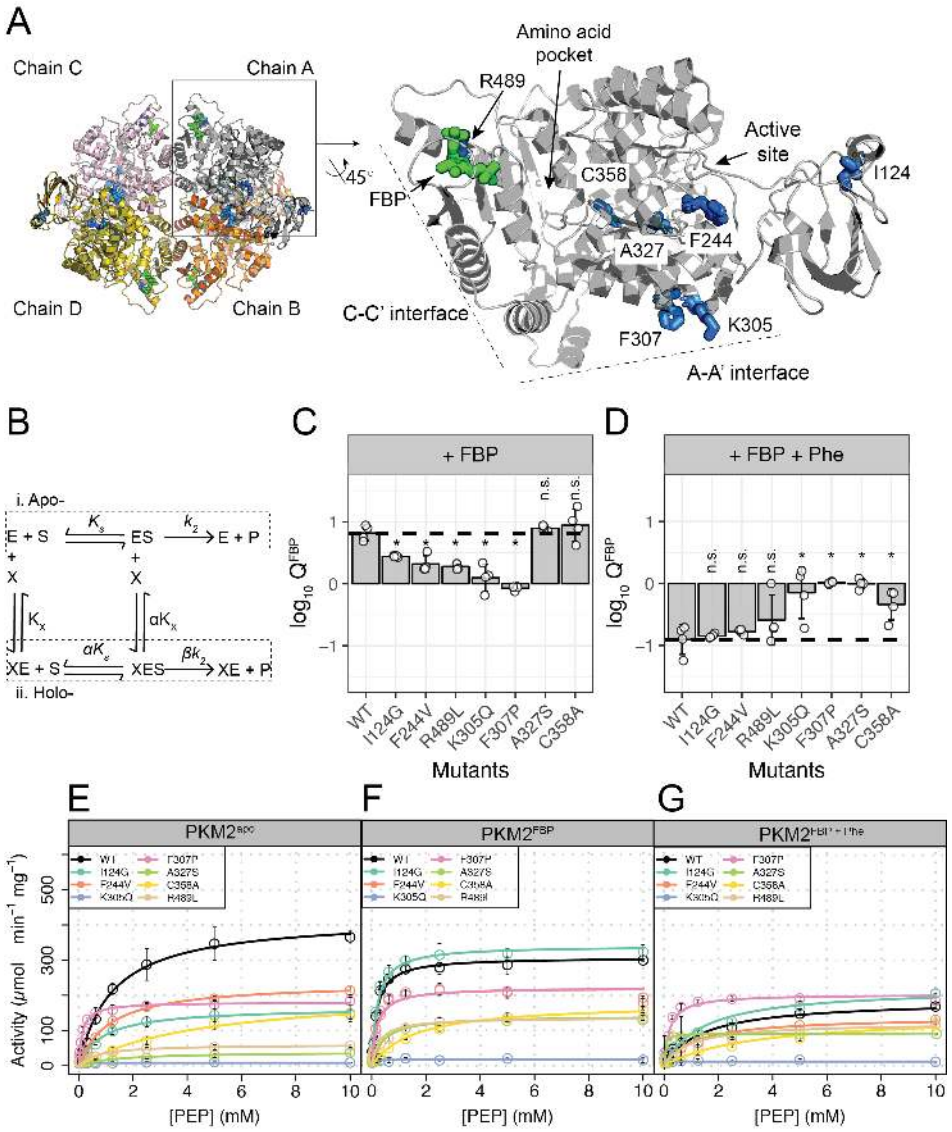

Figure 6.6: ***The AlloHubMuts perturb either FBP-active site coupling or FBP-Phe coupling.*** ***(A)*** *A structural schematic showing the locations of the AlloHubMuts on PKM2 protomer.* ***(B)*** *A single-substrate-single-modifier scheme was used to quantify the coupling co-efficient (Q) between FBP and substrate binding and between FBP and Phe and substrate binding.* ***(C)*** *The allosteric response of PKM2(WT) and AlloHubMuts to FBP, quantified by the allosteric coefficient, which denotes the change of the K_M^{PEP} upon addition of saturating concentrations of FBP (see Methods Section 2.3.5). The Q-coefficient for wild type PKM2 (WT) is shown as a dotted line for comparison. Each of the Q-coefficients of the AlloHubMuts were statistically compared to PKM2(WT) using a Wilcoxon ranked-sum test ($n = 4$); a p-value < 0.05 was deemed significant (denoted by an asterisk); n.s.: not significant.* ***(D)*** *The magnitude of allosteric inhibition by Phe, in the presence of FBP, determined for PKM2(WT) and AlloHubMuts, quantified by the allosteric co-efficient Q as in (B). Raw initial velocity rate curves are shown in the absence of added ligands in* ***(E)****, in the presence of saturating FBP* ***(F)*** *and in the presence of saturating FBP and 400 μM Phe* ***(G)****.*

AlloHubM	Ligand	K_M^{PEP} (mM)	k_{cat} (s^{-1})	k_{cat}/K_M^{PEP} ($s^{-1}\cdot$ mM^{-1})
I124G	Apo	1.07 ± 0.13	190.27 ± 7.31	177.82 ± 56.23
	FBP	0.29 ± 0.04	307.10 ± 6.11	1058.97 ± 152.75
	FBP+Phe	1.44 ± 0.34	221.92 ± 17.26	153.98 ± 37.88
F244V	Apo	1.14 ± 0.11	237.44 ± 7.36	210.44 ± 27.15
	FBP	0.55 ± 0.07	279.62 ± 9.93	522.42 ± 87.51
	FBP+Phe	1.15 ± 0.30	222.51 ± 18.29	195.87 ± 55.46
K305Q	Apo	0.01 ± 0.01	8.06 ± 0.57	790.01 ± 605.56
	FBP	0.01 ± 0.04	8.40 ± 0.43	1038.01 ± 535.50
	FBP+Phe	0.04 ± 0.01	12.21 ± 1.40	408.37 ± 136.10
F307P	Apo	0.13 ± 0.01	180.47 ± 3.05	1413.87 ± 133.12
	FBP	0.15 ± 0.02	227.20 ± 5.02	1508.53 ± 190.71
	FBP+Phe	0.21 ± 0.03	328.71 ± 12.37	1568.32 ± 268.58
A327S	Apo	1.37 ± 0.42	31.8 ± 2.43	23.21 ± 5.79
	FBP	0.17 ± 0.02	119.01 ± 6.86	700.06 ± 343.00
	FBP+Phe	0.15 ± 0.02	100.19 ± 5.32	667.93 ± 266.00
C358A	Apo	4.71 ± 0.99	214.73 ± 21.40	45.59 ± 21.62
	FBP	0.58 ± 0.18	191.13 ± 16.22	329.53 ± 90.11
	FBP+Phe	1.25 ± 0.07	199.01 ± 21.87	159.20 ± 31.24
R489L	Apo	0.69 ± 0.19	60.05 ± 4.91	89.38 ± 36.93
	FBP	0.36 ± 0.10	112.25 ± 7.60	317.20 ± 125.71
	FBP+Phe	1.44 ± 0.40	132.99 ± 15.74	158.19 ± 122.56

Table 6.3: ***Steady-state enzyme kinetic parameters of the AlloHubMuts.*** *The enzyme activity of the AlloHubMuts were measured over a range of phosphoenolpyruvate concentrations between 0 mM and 10 mM in the absence of added ligands, in the presence of saturating concentrations of FBP and in the presence of FBP and 400 μM (physiological concentrations) Phe. A protein concentration of 5 nM was used for all measurements. Activity measurements were repeated four times; the mean and standard deviations are shown.*

6.3.3 Native spectra of AlloHubMut uncouples oligomerisation from allosteric activation

Native spectra of PKM2(WT), in Chapter 4, showed that FBP-induced allosteric activation is accompanied by stabilisation of the high-substrate affinity tetrameric state. This observation led to the hypothesis that tetramerisation and enzyme activation are coupled. It would be expected, therefore, that a mutant which prevents tetramer formation would reveal a reduced (or no) allosteric response to FBP, prompting the lack of oligomerisation as a possible explanation as to why some of the AlloHubMuts had a reduced coupling between the FBP pocket and the active site (Sections 6.3.1 and 6.3.2).

To this end, native spectra of the AlloHubMuts were acquired in the absence of any added ligands (**Fig. 6.7 A**), and following pre-incubation with FBP at a ratio of 2:1 (ligand-to-protein) (**Fig. 6.7 B**). A327S and C358A both formed of monomers, dimers and tetramers, at an approximate ratio of 1:3:6 (**Fig. 6.7 C**). The addition of FBP to A327S and C358A resulted in a complete conversion of lower-order oligomers into tetramers, similar to PKM2(WT) (**Fig. 6.7 C**). In contrast, F244V and R489L formed monomer-dimer-tetramer equilibria, both at approximate ratios of 1:3:6, which was unperturbed following the addition of FBP (**Fig. 6.7 C**). The oligomeric response of A327S and C358A and the lack of response of F244V and R489L fit with the previous observation that A327S and C358A maintain an intact allosteric coupling between the FBP pocket and the active site, and that F244V and R489L have a significantly reduced enzymatic response to FBP (Section 6.3.1).

PKM2(I124G) formed a monomer-dimer-tetramer mixture, in the absence of added ligands, at an ratio equivalent to that of the wild-type (**Fig. 6.7 C**). The addition of FBP to I124G resulted in an apparent tetramerisation, despite the observation that this mutant has a significantly reduced enzymatic response to FBP binding. Conversely, K305Q exclusively formed monomers in the absence of FBP, likely due to the removal of a charged interaction between K305 and E384 at the A-A' interface (**Fig. 6.3 H**). Nevertheless, the addition of FBP

to K305Q resulted in the formation of a significant fraction of tetramers, with a monomer-dimer-tetramer distribution of approximately 4:1:5 (**Fig. 6.7**). Taken together, the finding that I124G and K305Q undergo an apparent tetramerisation upon FBP addition, despite the negated coupling between the FBP pocket and the active site, provided evidence to suggest that PKM2 allosteric activation is uncoupled from oligomerisation.

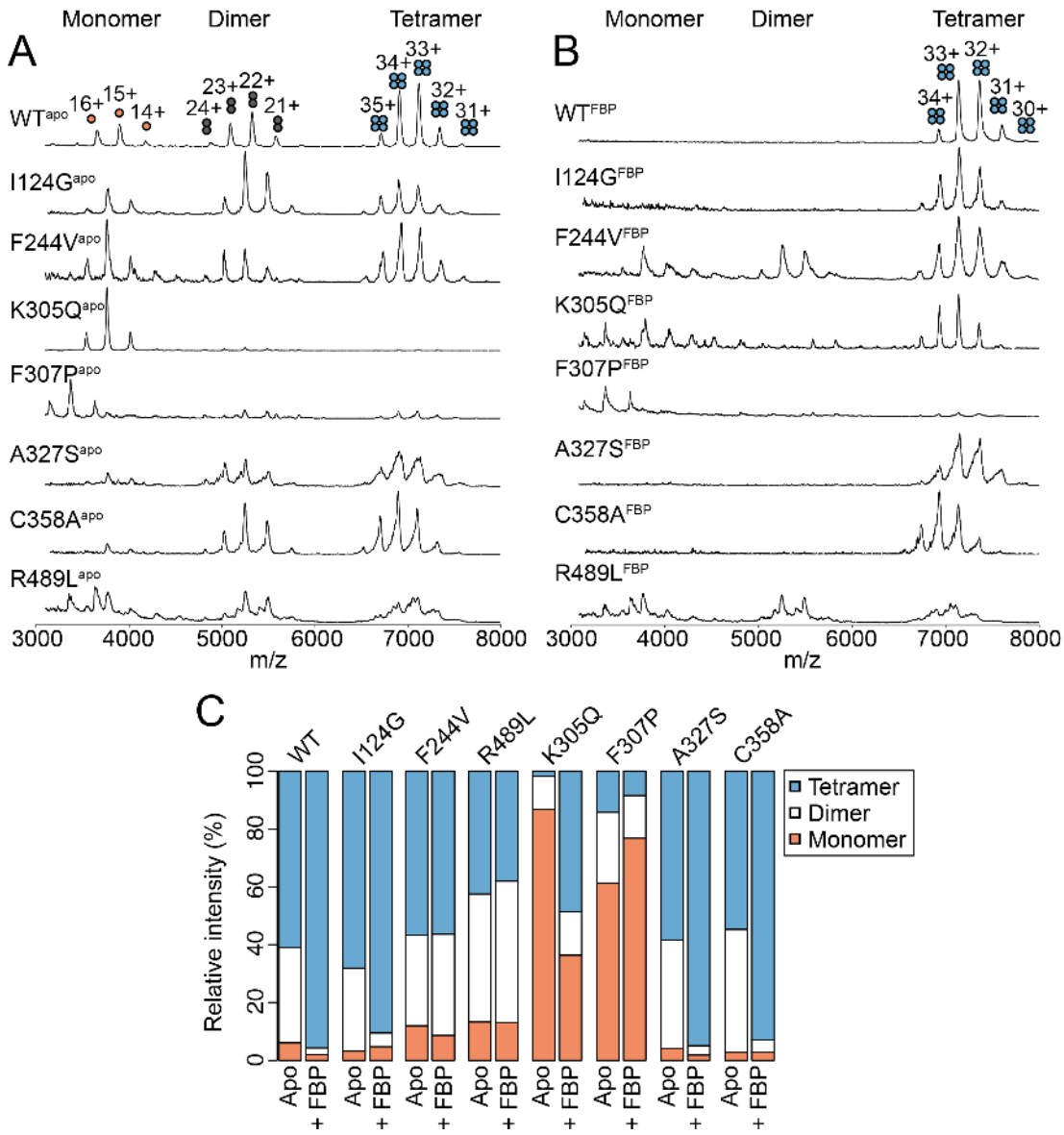

Figure 6.7: ***Native spectra of the AlloHubMuts reveal a complex relationship between allosteric activation and oligomerisation.*** *Nano-electrospray ionisation mass spectrometry was used to acquire native spectra of* ***(A)*** *PKM2(WT) and the AlloHubMuts, individually in the absence of added ligands and* ***(B)*** *in the presence of FBP. PKM2 variants were pre-incubated with FBP at a ratio of (2:1; protein-to-ligand).* ***(C)*** *The oligomeric state distribution was quantified for each of the AlloHubMuts and presented as stacked bars showing the relative intensities of monomers (orange), dimers (white) and tetramers (blue).*

6.4 Conclusion

In summary, evaluation of the allosteric properties of AlloHubMuts demonstrated that a novel computational method *AlloHubMat* identified residues involved in FBP-induced allostery. Two of the AlloHubMuts were found not to response to FBP *per se*, but rather to attenuate Phe-induced disruption of allosteric activation by FBP, revealing two residues (A327 and C358) that mediate a functional cross-talk between allosteric networks elicited from distinct ligand binding pockets on PKM2.

Chapter 7

Discussion

Cellular life exists under conditions where the concentrations of nutrients, growth factors and oxygen are in constant flux. The dynamic instability of extracellular conditions requires that biological catalysts are highly efficient so that essential chemical compounds are generated at sufficient concentrations from the available extracellular nutrient pool. Moreover, the rate of chemical synthesis must be tightly controlled to maintain cellular homeostasis in response to signalling cues. Allostery is an important feature of enzyme regulation, which is potent and highly specific, and contributes towards maintaining organised flux through cellular pathways. Despite the ubiquity and demonstrated importance of allosteric regulation, an understanding of its mechanisms at the molecular level is limited, as is the curration of proteome-wide data on allosteric proteins.

A detailed description of the regulatory mechanisms of PKM2 is relevant for understanding metabolic reprogramming in disease settings such as cancer, where it has been shown that inhibition of PKM2 activity can promote cancer cell growth by facilitating the provision of glucose-derived carbons for anabolic processes such as *de novo* lipogenesis[17], serine biosynthesis[19], nucleotide synthesis[17,44], and the production of reducing equivalents such as glutathione[40]. The discovery of a class of small molecules that activate PKM2 catalysis[47–49],

and the finding that these small molecule activators alleviate the progression of xenograft breast tumours[17], support the hypothesis that exogenous regulation of PKM2 activity acts on a metabolic vulnerability of cancer cells[25,31,32,34]. Given the cancer-associated role of PKM2 regulation, knowledge of the molecular processes by which ligand binding incites changes to enzyme activity are of therapeutic interest.

In addition to feed-forward activation by fructose 1,6-bisphosphate (FBP), PKM2 is regulated by a number of other endogenous ligands, including the amino acids L-serine (Ser; an allosteric activator)[64] and L-phenylalanine (Phe; an allosteric inhibitor)[168]. Investigations into PKM2 regulation, thus far, have focused on the effects of individual ligands *per se* on the structure[17,20,55,63], function[17,19,23] and dynamics[62,66,74] of PKM2. It is unclear whether ligands that concurrently bind to distinct allosteric pockets elicit in functionally independent effects, or whether the binding of multiple allosteric ligands with opposing functional signals produces a synergistic response.

Here I offer evidence that PKM2 regulation constitutes a functional cross-talk between the effects of FBP and amino acids, which act in a combinatorial manner to modulate the allosteric transition of the tetrameric form of the protein between the inactive and active states. Using a novel method for predicting allosteric hub residues, *AlloHubMat*, a network of protein residues is identified that couple FBP binding with enhanced substrate affinity. Moreover, two residues (A327 and C358) are identified that facilitate the cross-talk between FBP and Phe.

7.1 PKM2 is concurrently regulated by multiple allosteric ligands in a range of cellular conditions

An investigation into PKM2 regulation was initiated by determining the binding affinity of PKM2 to its endogenous regulators. An apparent dissociation constant in the nano-molar range [$K_D^{FBP} = (21.4 \pm 9.0)$ nM] was estimated from measurements of FBP binding to PKM2 (Section 3.2.1). Tight binding between FBP and PKM2 resulted in purified recombinant PKM2 retaining substoichiometric amounts of *E. coli*-derived FBP throughout the purification, despite extensive dialysis (Section 3.2.1). Amounts of co-purified FBP were quantified in all purifications of PKM2 and found that as much as 75 % of the protein was pre-bound to the ligand (Section 3.2.1). Previous studies have reported co-purification of FBP with PKM2, suggesting a similarly high binding affinity as that reported here[22,23,63]. Our estimate of the binding affinity is in approximate agreement with that of Yan *et al.* (2016)[62]: $K_D^{FBP} = 210$ nM; and Gavriilidou *et al.* (2018)[63]: $K_D^{FBP} = 910$ nM. Neither publication, however, provide estimates of the amounts of FBP in the starting material of purified PKM2 and it is therefore unlikely that these studies account for the effect of ligand co-purification on the starting free protein concentration.

The consequence of tight binding between FBP and PKM2 has been overlooked, thus far, as it raises questions regarding the biological role of FBP as an allosteric regulator of PKM2. To this end, intracellular concentrations of FBP were measured under fully-fed and glucose-deprived culture conditions in three human cancer cell lines (Section 3.5.1). Our findings revealed that intracellular FBP concentrations far exceeded the concentration of ligand required to achieve full saturation of PKM2. Consequently, under steady-state growth conditions, a significant fraction of PKM2 is predicted to be bound to FBP, even in the background of other regulatory stimuli such as post-translational-modifications, that influence FBP binding[23,82] (Section 3.5.3).

A protein concentration of 5 μM was used to measure the fluorescence spectrum of PKM2 and thereby monitor FBP binding (Section 3.2.1), which was approximately 240-fold greater than the calculated FBP binding affinity. Ideally, the concentrations of protein and ligand used for binding experiments should be in the same range as the measured K_D. Owing to the low quantum yield of tryptophan in intrinsic PKM2 fluorescence measurements, it was necessary to use PKM2 concentrations in the low μM range in order to obtain measurements with acceptable signal-to-noise ratios. Furthermore, PKM2 is a notoriously 'sticky' protein with a tendency to bind to the walls of the tube. While for many of the assays it was possible to avoid this (use of coated tubes, minimal sample mixing) diluting PKM2 to low concentrations for fluorescence measurements leads to significant loss of protein that makes such conditions unsuitable.

With these considerations in mind, the question is: can we derive meaningful K_D values from our binding data? Simulations of ligand-receptor binding, and fitting of these data with the quadratic expression in Equ. 3.4, found that as the free protein concentration increases, the % error in calculating the K_D also increases and reaches 130 - 175 % at protein concentrations 240-fold in excess of the binding affinity (similar to our experimental conditions in measuring FBP-PKM2 binding). This observation also reflects the large experiemental error reported for the K_D of FBP for PKM2 (48 %). Nevertheless, in spite of the large error associated with this measurement, due to the experimental limitations detailed above, we are still able to determine a range and an upper limit for the K_D values we measure. The upper-limit of the FBP-PKM2 binding affinity, taking into account the theoretical percentage error, corresponds to 58 nM. Importantly, this upper limit is between 500- and 4000-fold lower that the measured intracellular concentration of FBP and therefore our assertion that PKM2 is saturated remains valid. This assertion is further supported by the persistence of FBP binding to purified PKM2 despite extensive dialysis during the process of purification. Furthermore, in addition to FBP binding to PKM2, we also determined the AC95, which indicates the concentration of FBP that results in 95 % PKM2 activation, which we show is 529 nM i.e. 610-fold lower than the

measured intracellular concentration (Section 3.2.3).

The notion that FBP is constitutively bound to PKM2 under a range of extracellular glucose conditions (0 mM - 11 mM) is supported by the finding that addition of exogenous FBP had little effect on PKM2 activity assayed in cell lysates, cultured both in fully-fed and glucose-starved conditions (Section 3.5.2). Moreover, exogenous addition of phenylalanine to HCT116 cell lysates resulted in a dose-dependent decrease in the k_{cat} (Section 3.5.2), rather than an increase in the K_M^{PEP} observed when purified recombinant PKM2 activity was assay upon addition of phenylalanine (Section 3.4.1). While seemly contradictory, Phe inhibition of recombinant purified PKM2 in the presence of saturating concentrations of FBP was found also to reduce the k_{cat}, revealing an alternate mechanism of inhibition by Phe, and possibly other inhibitory amino acids, in the presence of constitutively bound FBP (Section 3.4.1), further supporting the hypothesis that, under the presently tested cell culture conditions, PKM2 is constitutively bound to FBP. Additionally, our results show that competition between activating (Ser) and inhibiting (Phe) amino acids determines the rate of product turnover of PKM2, while constitutive FBP binding maintains the protein in a high-substrate affinity state (Section 3.4.1).

Other amino acids L-alanine (Ala), L-tryptophan (Trp) and L-cysteine (Cys) have previous been shown to inhibit PKM2 activity *per se* in the same way as Phe[20]. Therefore, competition between activating and inhibitory amino acids may provide a potent mechanism of feed-forward activation and feed-back inhibition of FBP-bound PKM2 activity (**Fig. 7.1**). Under conditions of high PKM2 activity, glucose-derived phosphoenolpyruvate is rapidly converted to pyruvate, expanding the intracellular pool size of pyruvate. Excess pyruvate can then be consumed by a number of reactions including reduction by lactate dehydrogenase to form lactate, or oxidation by pyruvate dehydrogenase to form acetyl-CoA. Alternatively pyruvate can be used as a substrate by one of the alanine aminotransferase enzymes. Alanine aminotransferase 1 (GPT)-catalysed transamination between pyruvate and glutamate forms

α-ketoglutarate and alanine (**Fig. 7.1**). Accumulation of the alanine pool would feed-back inhibit FBP-bound PKM2 ($PKM2^{FBP}$), likely by the same hyperbolic-mixed mechanism observed for phenylalanine. Sustained allosteric inhibition of $PKM2^{FBP}$ would result in the accumulation of glycolytic intermediates, leading to utilisation of 3-phosphoglycerate for *de novo* serine biosynthesis. Subsequent accumulation of serine would feed-forward activate PKM2 (**Fig. 7.1**), as shown by Chaneton *et al.* (2012)[19]. The findings presented herein, which support the idea that PKM2 is constitutively saturated with the activator FBP, suggest that serine-induced activation does not act to increase the substrate affinity of PKM2. Rather, high concentrations of serine, relative to the concentrations of alanine and other inhibitory amino acids, would act to out-compete the inhibitory amino acids from the amino acid binding pocket on PKM2, thereby alleviating the negative effect on the k_{cat}. Given that Phe, Ser, Ala, His and Trp bind to a common allosteric pocket[20], it is likely that their relative, rather than absolute, concentrations determine the activity of FBP-bound PKM2 (**Fig. 7.1**). Whether intracellular flux through lower glycolysis is determined by the ratio between activating-to-inhibiting amino acid ligands is beyond the scope of this thesis, and warrants further investigation.

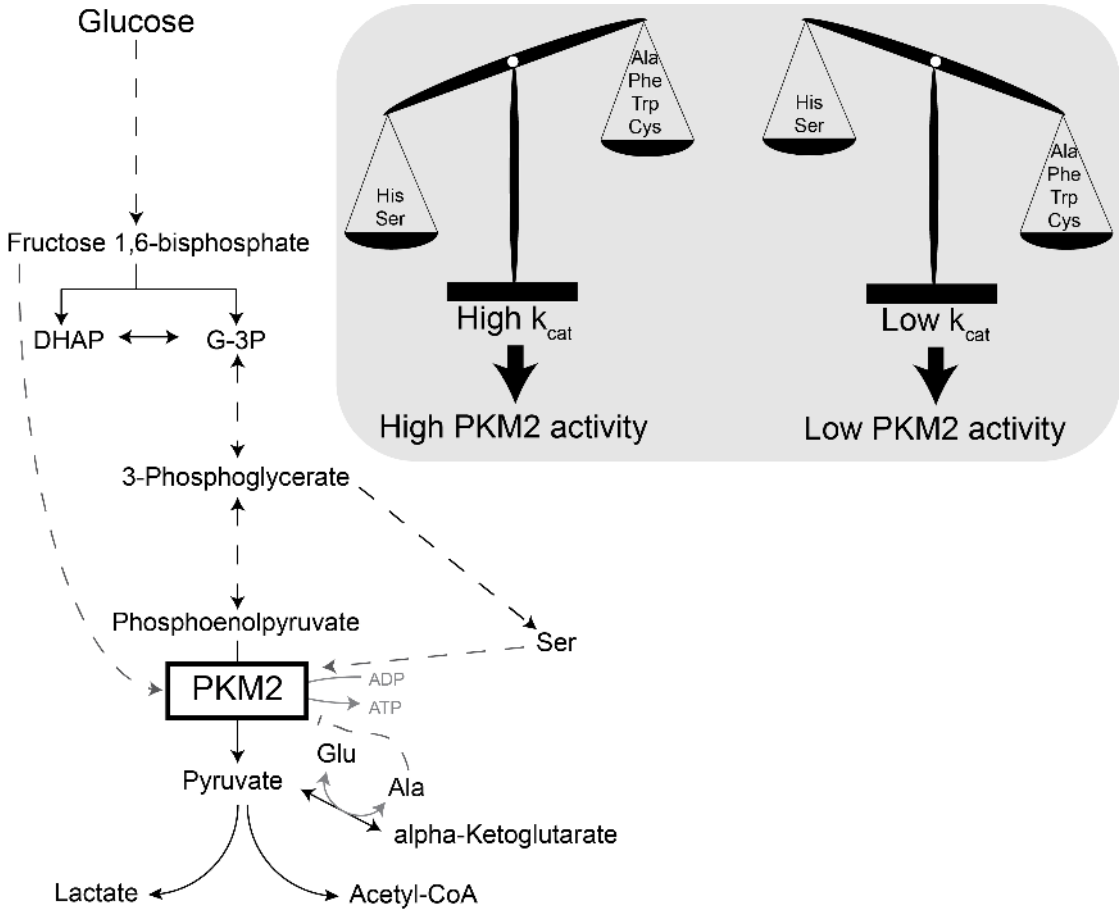

Figure 7.1: ***Proposed model of PKM2 regulation by amino acids.*** *Work presented herein suggests that FBP is constitutively bound to PKM2, thereby locking the enzyme in a high-substrate affinity state. PKM2 enzyme activity can subsequently be regulated by activating and inhibitory amino acids which collectively compete for a single binding site. Therefore, high concentrations of inhibitory amino acids relative to activating amino acids, would act to reduce the rate of product turnover through a hyperbolic-mixed allosteric mechanism. Conversely, high concentrations of activating amino acids, relative to the pool size of inhibitory amino acids, would result in a high k_{cat} by alleviating the effect of inhibitory amino acids.*

7.2 FBP and Phe regulate PKM2 through a functional cross-talk

The proposed model of PKM2 regulation outlined in Section 7.1 is supported by the finding that, while FBP and Phe bind to spatially distinct pockets on PKM2, both ligands influence the mode of action of the other without reciprocal effects on their binding affinities. The ability of Phe to perturb FBP-induced activation, suggests a functional cross-talk between the allosteric mechanisms of FBP and Phe. Previous studies have suggested that PKM2 enzyme activity is correlated with oligomerisation[17,20,22,59,63,64,70,168]. Therefore, to explore whether the combined effects of concurrent FBP and Phe regulation could be explained by changes to the oligomeric state of the protein, native mass spectrometry (MS) was used as a means to study the oligomeric structure and dynamics of PKM2 in the gas phase (Chapter 4).

Native spectra of PKM2 revealed that the protein adopts a mixed population of monomers, dimers and tetramers (Section 4.2.1). FBP addition was found to promote tetramerisation (Section 4.2.3), in agreement with a recent native MS study[63]. An integrative approach using ion-mobility MS and *in vacuo* MD simulations found that the dimers in the native spectra were formed stably about the A-A' interface (Section 4.2.2), which implies that FBP binding induces dimer-dimer association about the C-C' interface. This is supported by a previous study showing that a G415R mutation prevents tetramerisation about the C-C' interface and is thereby insensitive to FBP-induced activation[62]. In contrast, Phe addition had the effect of preferentially stabilising the dimeric species of PKM2 (Section 4.3.1). Our results thus support the concept that FBP-induced activation stabilises the tetrameric species, while Phe-induced inhibition promotes dimeric PKM2.

The finding herein that Phe binding *per se* destabilises tetramers is in agreement with previous studies[70,168], and contests the findings of two recent reports suggesting that Phe, and several other amino acids, promote the formation of an inactive tense *T-state* tetramer[20,22]. Critically,

our observation that concurrent addition of Phe and FBP synergistically promotes the formation of tetramers (Section 4.3.2) raises the possibility that Phe-induced stabilisation of the inactive tetramer reported by Morgan *et al.* (2013)[22] and Yuan *et al.* (2018)[20] is confounded by the presence of residually co-purified FBP. Indeed Morgan *et al.* (2013)[22] generated a mutant variant R489A in order to "[abolish] FBP binding and [prevent] contamination by FBP, which is commonly bound during purification of the protein from Escherichia coli culture." Nevertheless, it is unclear whether partial FBP occupancy is accounted for in these studies, as the amounts of co-purified FBP in preparations of recombinant purified PKM2 are not reported. It is notable, therefore, that Morgan *et al.* (2013)[22] chose to test oligomerisation upon Phe addition with PKM2(WT) (partially FBP-bound) rather than with PKM2(R489A) (fully apo). Based on the findings presented herein, one might speculate that the stabilisation of PKM2 tetramers by Phe observed by Morgan *et al.* (2013)[22] and Yuan *et al.* (2018)[20] could be attributed to significant amounts of residually co-purified FBP from *E. coli*, which we found to be as high as 78 % of the concentration of purified PKM2 (Section 3.2.1).

In addition to allosteric effects on PKM2 oligomerisation, several studies have invoked inter-domain structural changes consequent to ligand binding[20,22,55,62,77]. To this end, ion mobility coupled to native mass spectrometry (IM-MS) measurements were used to compute the rotationally-averaged collision cross section ($^{DT}CCS_{He}$) of PKM2 as a measure of the conformational heterogeneity of the protein. FBP binding resulted in a subtle shift in the $^{DT}CCS_{He}$ distribution, favouring a more extended form of the protein (Section 4.2.5). This was partially reversed by the simultaneous addition of both FBP and Phe, favouring a more apo-like $^{DT}CCS_{He}$ distribution (Section 4.3.4). The apparent ligand-induced changes in PKM2 ion mobility may indicate a conformational shift between the T-state (tense) and the R-state (relaxed) tetramer, as described previously by Morgan *et al.* (2013)[22] and Yuan *et al.* (2018)[20]. Although informative, it is difficult to relate changes in ion mobility with specific structural events. Therefore, to study the process of ligand-induced structural rearrangement, we turned to an *in silico* thermodynamic model of PKM2 (Chapter 5).

MD simulations of tetrameric PKM2 found that allosteric activators FBP and Ser induced a closure of the B-domain cap over the active site pocket, whereas simulations of PKM2$^{FBP+Phe}$ and PKM2^{Apo} tetramers revealed an opening of the B-domain cap (Section 5.2.5). The concept of PKM2 ligand-dependent B-domain cap dynamics is supported by high B-factors calculated fro atoms in the B-domain of a crystal structure previously published by Dombrauckas *et al.* (2005)[55]. This hypothesis is supported by molecular dynamics simulations published by Gehrig *et al.* (2017)[77] and by Naithani *et al.* (2015)[74], both describing an FBP-induced closure of the B-domain. Nevertheless, high-resolution evidence of the simulated phenomena remains elusive. Small-angle X-ray scattering measurements of PKM2 tetramers by Yan *et al.* (2016)[62] observed a small decrease in the radius of gyration upon addition of FBP from 43 Åto 41 Å. The authors attributed no significance to the observed FBP-induced change, stating that "binding of FBP did not significantly alter the R_g of PKM2(WT)"[62]. In retrospect, however, it is tempting to speculate that this small change in the R_g may reflect the B-domain cap closure in the T- to R-state tetramer transition. This could be addressed experimentally using methyl-TROSY (transverse relaxation optimised spectroscopy) NMR spectroscopy, which enables the study of the conformational dynamics of large biomolecular systems[179].

7.3 AlloHubMat reveals residues that mediate the cross-talk between FBP- and Phe-induced allosteric regulation

To identify residues that mediate allosteric regulation of PKM2, the conformational dynamics of PKM2 tetramers upon FBP addition were analysed using molecular dynamics simulations of PKM2^{apo} and PKM2^{FBP} tetramers. Building on a previous approach by Pandini *et al.* (2012)[123], the mutual information between sampled conformational states from MD trajectories encoded in a coarse-grained representation within the framework of the M32K25 structural alphabet[125]. Proteins have been shown both experimentally[117,180,181] and computationally[147,182] to sample distinct conformational sub-states. This sampling of phase space is relevant for allosteric regulation of enzyme activity[87,183–185], leading to the prevailing view that protein dynamics and allosteric regulation are ensemble phenomena[86]. Therefore, explicitly identifying allosteric signals that are representative of the ensemble of protein sub-states is critical. To this end, a novel computational framework, named AlloHubMat (**Allo**steric **Hub** prediction using **Mat**rices that capture allosteric coupling) was developed to predict allosteric hub fragments from the network of dynamic correlated motions, based on explicitly identified conformational sub-states from multiple MD trajectories and obtain an ensemble-averaged mutual information network (Section 5.3).

At its core, AlloHubMat uses the M32K25 structural alphabet (SA) as a low-dimensional representation of the torsional space accessible to proteins[125]. Each fragment in the alphabet contains two angles and a torsion angle, and is partitioned into 25 states forming a discrete-state model of protein structure. Previous SAs have used a similar fragment-based representation in the form of a set of Cartesian coordinates (MSM2000)[186], or a vectorial description of consecutive Cα atoms (CGT2004)[187]. The M32K25 SA has the advantage, compared to other SAs, of being derived from a density-based approach[188], so that the most dominant conformations are highly populated with multiple states[125]. Therefore, even relatively subtle changes in the protein backbone are captured by changes between states within the M32K25 model. Allo-

HubMat overcomes limitations of previous approaches[120,123] that do not account for how distal correlated motions may change in direction and magnitude as the protein samples distinct substates over the course of an MD simulation, and predicts networks of allosteric residues from MD simulations using a consistent numerical framework to measure time-dependent correlated motions. The approach used by AlloHubMat facilitates the extraction of consensus allosteric networks from replicate MD simulations of a protein in a given liganded state, and the comparison of the consensus networks between simulations of different liganded states.

Analysis of MD simulations of tetrameric PKM2 with AlloMatHub identified a number of candidate fragments as allosteric hub fragments (AlloHubFrags; Section 5.3.3). Starting from a selection of allosteric hub fragments (AlloHubFrags), single-point mutant variants (**Fig. 7.2 A**) were designed and tested for their enzymatic response to FBP binding. Mutagenesis of several of the hubs (I124G, F244V, K305Q, F307P and R489L) disrupted FBP-induced activation (Section 6.3.1), demonstrating the utility of AlloHubMat for identifying *bona fide* mediators of allostery. Of the five mutants that abrogated allosteric activation of PKM2, three mutants (F244V, F307P and R489L) disrupted the process of oligomerisation (Section 6.3.3), suggesting that these AlloHubs mediate the FBP-induced monomer/dimer to tetramer transition (**Fig. 7.2 B**). Conversely, AlloHubMuts I124G and K305Q maintained the ability to tetramerise upon FBP addition (Section 6.3.3), though their allosteric activation was perturbed (Section 6.3.1). Additional mutants of residues A327 and C358 preserved the allosteric coupling between FBP binding and enzyme activation, though prevented the inhibitory effect of Phe on PKM2^{FBP} activity (**Fig. 7.2 B**; Section 6.3.1). The finding that A327S and C358A maintained wild type-like FBP activation but perturbed Phe inhibition indicated a role for these two residues in mediating a cross-talk between the allosteric mechanisms of Phe and FBP, which enables the inhibitory amino acid to inhibit the effect of FBP. C358 has been previously identified as an oxidation substrate of reactive oxygen species, whereby PKM2 can be post-translationally modified resulting in enzyme inhibition[40]. None of the AlloHubMuts fell within residues 389-429 that differ between PKM2 (allosterically regulated) and the consti-

tutively active splice-isoform (PKM1), suggesting that residues that confer differences in the allosteric properties of the two isoforms are dispersed throughout the protein structure. Zhong *et al.* (2017)[66] recently reported a synergistic mechanism whereby adenosine monophosphate (AMP) and glucose-6-phosphate (G6P) activate *M. tuberculosis* pyruvate kinase. AMP binds to the equivalent of the FBP binding pocket, whereas G6P binds to a different pocket that is also distinct from the equivalent of the amino acid binding pocket on human PKM2. The synergistic mechanism between G6P and AMP is somewhat similar to the one described between FBP and Phe, here. Therefore, it is tempting to speculate that allosteric synergism upon concurrent binding of different ligands is a common mechanism of enzyme regulation.

Evidence presented in Chaters 5 and 6 suggests that AlloHubMat is able to identify residues involved in the allosteric mechanism of an enzyme. Nevertheless, a recent large-scale alanine scanning mutagenesis study of liver pyruvate kinase (PKL) approximated that more than 30 % of the protein residues are involved in its allosteric regulation[76]. Consequently, it can be argued that a statistical validation of AlloHubMat should focus on measuring the effect of mutating non-hub residues as negative controls for assessing the precision of the method. A comprehensive validation of the method was beyond the scope of this Thesis. Moreover, the mechanism enzyme regulation is likely case-dependent; for some proteins configurational entropy is the main driver of allostery[110,111,173–175], and for others enthapic motions dominate the allosteric transition[113,167,179,189–191]. This reasoning would conclude that a case-by-case validation of methods to predict allosteric regulation is required, which largely defeats the purpose of a predictive computational method. Moreover, high-throughput analyses of protein allostery using site-directed mutagenesis is compounded by the challenging task of purifying a very large number of mutant variants and investigating whether the introduced chemical perturbations abrogate coupling between distal sites, oligomerisation, or both. To this end, proteome-wide curation of data resulting from the experimental characterisation of enzyme mutants would address the outstanding challenge of benchmarking AlloHubMut and other computational methods for predicting allosteric hubs.

The spatial resolution of AlloHubMat is currently limited to fragments of four successive Cα atoms. In Chapter 5, single-point mutants were designed from four-residue candidate hubs, based on an empirical analysis of the per-residue conservation within the identified hubs. In addition to identifying hub fragments, it would be desirable to explicitly identify the residue(s) within a hub fragment that contribute most to ochestrating the allosteric mechanism. Given that neighbouring fragments were found to be correlated, it was not possible to derive the per-residue mutual information using the chain rule for mutual information (Section 5.3.3). Future developments could improve the resolution of AlloHubMat by incorporating other sequence-based[192,193] and statistical mechanical[194–196] methods, to build a combined predictive score. Sequence-based methods to measure residue co-evolution of distal residues[197] have been highly successful in predicting allosteric pathways in a number of cases[198,199], and would add chemical and evolutionary information that is lost in our coarse-grained approximation of protein dynamics.

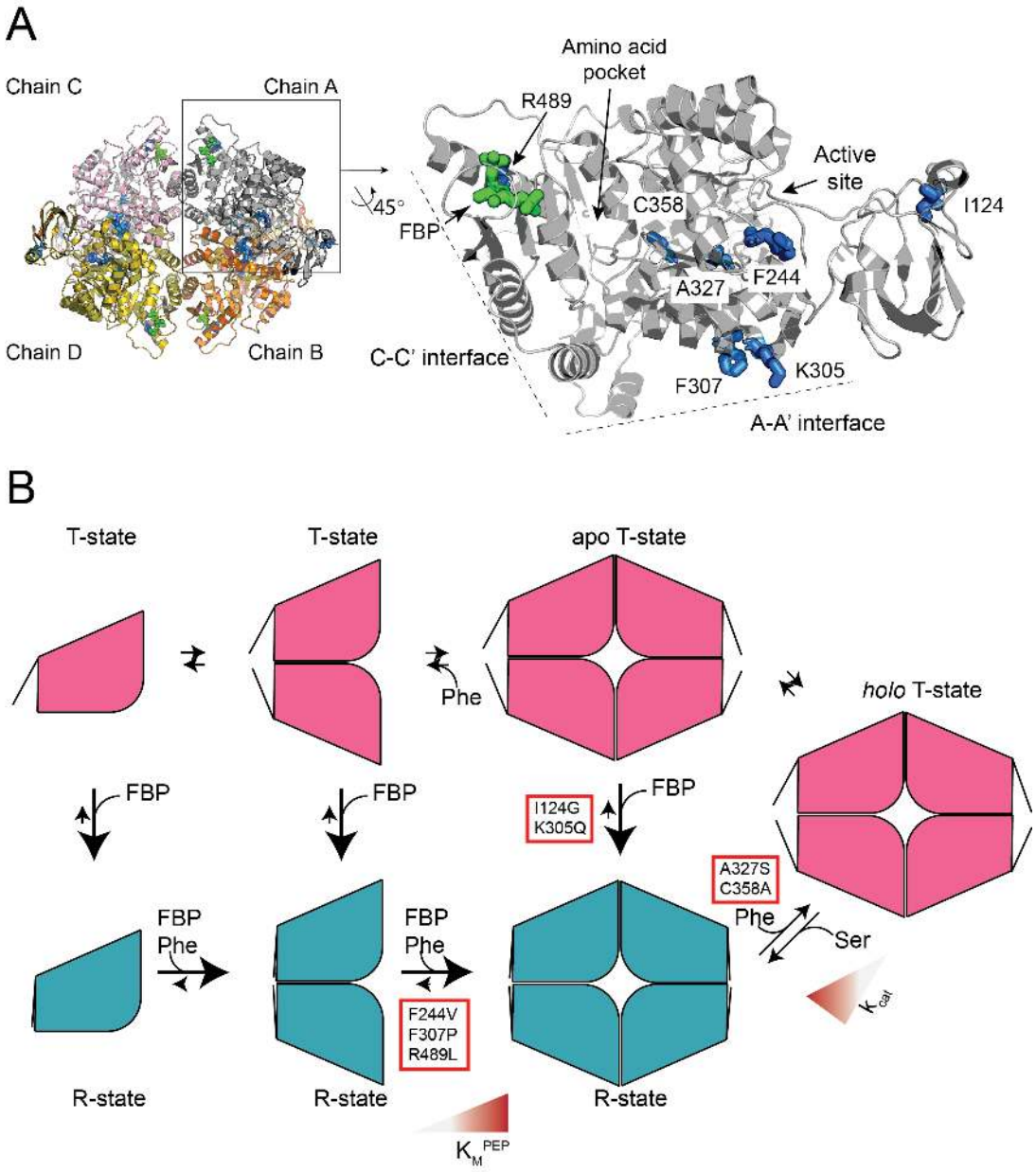

Figure 7.2: ***The location and function of the AlloHubMuts.*** ***(A)*** *A structural schematic of PKM2 showing the location the seven AlloHub mutations (AlloHubMuts) on the protomer. The FBP molecule is shown in green and the locations of the amino acid binding pocket and the active site are annotated.* ***(B)*** *FBP-induced tetramerisation is propagated by a network of residues involving F244, F307 and R489. An apo T-state tetramer is populated in the ensemble of available states, to which FBP can bind, inducing a transition to the R-state tetramer. This T-state to R-state transition is accompanied by subtle conformational changes, likely involving the closure of the B-domain cap over the active-site pocket, which is correlated with an increased substrate binding affinity. The R- to T-state transition is mediated by I124 and K305, and mutations at these positions perturb the formation of a high substrate-affinity R-state tetramer.*

7.4 Conclusion

Allosteric regulation of PKM2 by FBP has long been viewed as a prototypical example of feed-forward regulation in biology, though the molecular mechanism of FBP activation remained largely elusive. PKM2 is additionally regulated by a plethora of endogenous ligands including several amino acids. It has been unclear, thus far, how PKM2 integrates the signals elicited by concurrently bound ligands, and the functional consequences of multi-ligand binding. To this end, the allosteric mechanism of PKM2 was investigated using an integrative computational and experimental approach.

Estimations of the fraction of intracellular PKM2 bound to its ligands found that that FBP concentrations far exceed that, which is required for full saturation of PKM2. Under steady-state growth conditions, a significant fraction of PKM2 is already bound to the activator FBP, even in the context of other regulatory cues, such as PTMs, that may influence ligand binding. Constitutive PKM2-FBP binding implicated FBP as an 'allosteric co-factor', rather than a reversible activator of PKM2 catalysis, raising questions regarding how PKM2 activity is regulated in the context of saturating FBP. Subsequent to FBP binding, amino acids Phe and Ser were found to reversibly modulate the maximal velocity of FBP-bound PKM2, suggesting an important role of amino acid regulation in lower glycolysis.

Simultaneous Phe and FBP binding was found to synergistically promote PKM2 tetramerisation, despite the two ligands revealing opposing effects on oligomerisation *per se*, revealing a functional cross-talk. To unravel the atomic basis by which allosteric coupling between distal sites is propagated within a protein, a novel computational method *AlloHubMat* was developed. Analysis of molecular dynamics simulations using AlloHubMat, and extensive experimental characterisation of candidate mutant variants, found that the process of FBP-induced activation is propagated by a network of residues involving I124, F244, K305, F307 and R489. Moreover, mutations of A327 and C358 were found to perturb the effect of Phe on

FBP-induced activation, thus revealing molecular details of the FBP-Phe cross-talk.

In summary, results herein reveal that PKM2 integrates multiple allosteric inputs to regulate enzyme activity. This phenomenon is analogous to multiple-input-single-output (MISO) controllers in control system engineering, which integrate multiple transmission signals (allosteric ligands) and relay a single signal to a receiver (enzyme activity). Beyond PKM2, it is likely that many more proteins in the cell possess the ability to bind to multiple regulatory ligands. Though whether a systems-control mechanism of integrating the inputs from multiple ligands is a general property of such proteins is unknown. Identifying allosteric hubs using AlloHubMat to design mutants that perturb the allosteric response of proteins to individual, or multiple, ligands presents a powerful toolbox for the study of both the mechanistic basis of allosteric signal integration as well as the functional consequences of combinatorial allosteric inputs on enzyme regulation.

Bibliography

[1] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–4 (2000).

[2] Oparin, A. I. *The Origin of Life: Translation with Annotations by Sergius Morgulis* (Dover Publications, 1965).

[3] Keller, M. A., Turchyn, A. V. & Ralser, M. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible archean ocean. *Mol Syst Biol* **10**, 725 (2014).

[4] Ralser, M. The RNA world and the origin of metabolic enzymes. *Biochem Soc Trans* **42**, 985–8 (2014).

[5] Konhauser, K. O. *et al.* Aerobic bacterial pyrite oxidation and acid rock drainage during the great oxidation event. *Nature* **478**, 369–73 (2011).

[6] Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* **39**, 99–123 (2004).

[7] Lincoln, T. A. & Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–32 (2009).

[8] Fusz, S., Eisenführ, A., Srivatsan, S. G., Heckel, A. & Famulok, M. A ribozyme for the aldol reaction. *Chem Biol* **12**, 941–50 (2005).

[9] Aisen, P. & Listowsky, I. Iron transport and storage proteins. *Annu Rev Biochem* **49**, 357–393 (1980).

[10] Tanner, L. B. *et al.* Four key steps control glycolytic flux in mammalian cells. *Cell Syst* (2018).

[11] Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).

[12] Novick, A. & Szilard, L. Experiments with the chemostat on the rates of amino acid synthesis in bacteria. *Proc Natl Acad Sci USA* **36**, 708–719 (1950).

[13] Goyal, S., Yuan, J., Chen, T., Rabinowitz, J. D. & Wingreen, N. S. Achieving optimal growth through product feedback inhibition in metabolism. *PLoS Comput Biol* **6**, 100–122 (2010).

[14] Chance, B., Hess, B. & Betz, A. DPNH oscillations in a cell-free extract of S. carlsbergensis. *Biochem Biophys Res Commun* **16**, 182–7 (1964).

[15] Das, J. & Busse, H. G. Analysis of the dynamics of relaxation type oscillation in glycolysis of yeast extracts. *Biophys J* **60**, 369–79 (1991).

[16] Hung, Y. P., Albeck, J. G., Tantama, M. & Yellen, G. Imaging cytosolic nadh-nad(+) redox state with a genetically encoded fluorescent biosensor. *Cell Metab* **14**, 545–54 (2011).

[17] Anastasiou, D. *et al.* Pyruvate Kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nat Chem Biol* **8**, 839–47 (2012).

[18] Sparmann, G., Schulz, J. & Hofmann, E. Effects of l-alanine and fructose (1,6-diphosphate) on pyruvate kinase from ehrlich ascites tumour cells. *FEBS Lett* **36**, 305–8 (1973).

[19] Chaneton, B. *et al.* Serine is a natural ligand and allosteric activator of pyruvate kinase M2. *Nature* **491**, 458–62 (2012).

[20] Yuan, M. *et al.* An allostatic mechanism for m2 pyruvate kinase as an amino-acid sensor. *Biochem J* **475**, 1821–1837 (2018).

[21] Keller, K. E., Tan, I. S. & Lee, Y.-S. SAICAR stimulates pyruvate kinase isoform M2

and promotes cancer cell survival in glucose-limited conditions. *Science* **338**, 1069–72 (2012).

[22] Morgan, H. P. *et al.* M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. *Proc Natl Acad Sci USA* **110**, 5881–6 (2013).

[23] Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–6 (2008).

[24] Kato, H., Fukuda, T., Parkison, C., McPhie, P. & Cheng, S. Y. Cytosolic thyroid hormone-binding protein is a monomer of pyruvate kinase. *Proc Natl Acad Sci USA* **86**, 7861–5 (1989).

[25] Dayton, T. L., Jacks, T. & Vander Heiden, M. G. PKM2, cancer metabolism, and the road ahead. *EMBO Rep* **17**, 1721–1730 (2016).

[26] Christofk, H. R. *et al.* The m2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–3 (2008).

[27] Lüftner, D. *et al.* Tumor type m2 pyruvate kinase expression in advanced breast cancer. *Anticancer Res* **20**, 5077–82 (2000).

[28] Lu, W. *et al.* Up-regulation of PKM2 promote malignancy and related to adverse prognostic risk factor in human gallbladder cancer. *Sci Rep* **6**, 26351 (2016).

[29] Chen, J. *et al.* Shikonin and its analogs inhibit cancer cell glycolysis by targeting tumor pyruvate kinase-m2. *Oncogene* **30**, 4297–306 (2011).

[30] Vander Heiden, M. G. *et al.* Identification of small molecule inhibitors of pyruvate kinase M2. *Biochem Pharmacol* **79**, 1118–24 (2010).

[31] Vander Heiden, M. G. *et al.* Metabolic pathway alterations that support cell proliferation. *Cold Spring Harb Symp Quant Biol* **76**, 325–34 (2011).

[32] Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–33 (2009).

[33] Mazurek, S. Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. *Int J Biochem Cell Biol* **43**, 969–80 (2011).

[34] Israelsen, W. J. & Vander Heiden, M. G. Pyruvate kinase: Function, regulation and role in cancer. *Semin Cell Dev Biol* **43**, 43–51 (2015).

[35] Jiang, L. & Deberardinis, R. J. Cancer metabolism: When more is less. *Nature* **489**, 511–2 (2012).

[36] Halliwell, B. Oxidative stress and cancer: have we moved forward? *Biochem J* **401**, 1–11 (2007).

[37] Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic Biol Med* **49**, 1603–16 (2010).

[38] Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–13 (2009).

[39] Wellen, K. E. & Thompson, C. B. Cellular metabolic stress: considering how cells respond to nutrient excess. *Mol Cell* **40**, 323–32 (2010).

[40] Anastasiou, D. *et al.* Inhibition of pyruvate kinase m2 by reactive oxygen species contributes to cellular antioxidant responses. *Science* **334**, 1278–83 (2011).

[41] Yang, M. & Vousden, K. H. Serine and one-carbon metabolism in cancer. *Nat Rev Cancer* **16**, 650–62 (2016).

[42] Maddocks, O. D. K. *et al.* Serine starvation induces stress and p53-dependent metabolic remodelling in cancer cells. *Nature* **493**, 542–6 (2013).

[43] Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–70 (2005).

[44] Kim, D. *et al.* SHMT2 drives glioma cell survival in ischaemia but imposes a dependence on glycine clearance. *Nature* **520**, 363–7 (2015).

[45] Israelsen, W. J. *et al.* PKM2 isoform-specific deletion reveals a differential requirement for pyruvate kinase in tumor cells. *Cell* **155**, 397–409 (2013).

[46] Lunt, S. Y. *et al.* Pyruvate kinase isoform expression alters nucleotide synthesis to impact cell proliferation. *Mol Cell* **57**, 95–107 (2015).

[47] Boxer, M. B. *et al.* Evaluation of substituted n,n'-diarylsulfonamides as activators of the tumor cell specific m2 isoform of pyruvate kinase. *J Med Chem* **53**, 1048–55 (2010).

[48] Jiang, J. K. *et al.* Evaluation of thieno[3,2-b]pyrrole[3,2-d]pyridazinones as activators of the tumor cell specific m2 isoform of pyruvate kinase. *Bioorg Med Chem Lett* **20**, 3387–93 (2010).

[49] Walsh, M. J. *et al.* 2-oxo-n-aryl-1,2,3,4-tetrahydroquinoline-6-sulfonamides as activators of the tumor cell specific m2 isoform of pyruvate kinase. *Bioorg Med Chem Lett* **21**, 6322–7 (2011).

[50] Kung, C. *et al.* Small molecule activation of PKM2 in cancer cells induces serine auxotrophy. *Chem Biol* **19**, 1187–98 (2012).

[51] Dayton, T. L. *et al.* Germline loss of PKM2 promotes metabolic distress and hepatocellular carcinoma. *Genes Dev* **30**, 1020–33 (2016).

[52] Morita, M. *et al.* Pkm1 confers metabolic advantages and promotes cell-autonomous tumor cell growth. *Cancer Cell* **33**, 355–367 (2018).

[53] Allen, A. E. & Locasale, J. W. Glucose metabolism in cancer: The saga of pyruvate kinase continues. *Cancer Cell* **33**, 337–339 (2018).

[54] Larsen, T. M., Benning, M. M., Rayment, I. & Reed, G. H. Structure of the bis(Mg^{2+})-ATP-oxalate complex of the rabbit muscle pyruvate kinase at 2.1 Å resolution: ATP binding over a barrel. *Biochemistry* **37**, 6247–55 (1998).

[55] Dombrauckas, J. D., Santarsiero, B. D. & Mesecar, A. D. Structural basis for tumor pyruvate kinase m2 allosteric regulation and catalysis. *Biochemistry* **44**, 9417–29 (2005).

[56] Guo, C. *et al.* Discovery of 2-((1h-benzo[d]imidazol-1-yl)methyl)-4h-pyrido[1,2-a]pyrimidin-4-ones as novel PKM2 activators. *Bioorg Med Chem Lett* **23**, 3358–63 (2013).

[57] Wang, P., Sun, C., Zhu, T. & Xu, Y. Structural insight into mechanisms for dynamic regulation of PKM2. *Protein Cell* **6**, 275–87 (2015).

[58] Srivastava, D., Razzaghi, M., Henzl, M. T. & Dey, M. Structural investigation of a dimeric variant of pyruvate kinase muscle isoform 2. *Biochemistry* **56**, 6517–6520 (2017).

[59] Ikeda, Y. & Noguchi, T. Allosteric regulation of pyruvate kinase m2 isozyme involves a cysteine residue in the intersubunit contact. *J Biol Chem* **273**, 12227–33 (1998).

[60] Yacovan, A. *et al.* 1-(sulfonyl)-5-(arylsulfonyl)indoline as activators of the tumor cell specific m2 isoform of pyruvate kinase. *Bioorg Med Chem Lett* **22**, 6460–8 (2012).

[61] Akhtar, K. *et al.* Differential behavior of missense mutations in the intersubunit contact domain of the human pyruvate kinase m2 isozyme. *J Biol Chem* **284**, 11971–81 (2009).

[62] Yan, M. *et al.* Succinyl-5-aminoimidazole-4-carboxamide-1-ribose 5'-phosphate (saicar) activates pyruvate kinase isoform m2 (PKM2) in its dimeric form. *Biochemistry* **55**, 4731–6 (2016).

[63] Gavriilidou, A. F. M. *et al.* Native mass spectrometry gives insight into the allosteric binding mechanism of m2 pyruvate kinase to fructose-1,6-bisphosphate. *Biochemistry* **57**, 1685–1689 (2018).

[64] Eigenbrodt, E., Leib, S., Krămer, W., Friis, R. R. & Schoner, W. Structural and kinetic differences between the m2 type pyruvate kinases from lung and various tumors. *Biomed Biochim Acta* **42**, S278–82 (1983).

[65] Ashizawa, K., McPhie, P., Lin, K. H. & Cheng, S. Y. An in vitro novel mechanism of regulating the activity of pyruvate kinase m2 by thyroid hormone and fructose 1,

6-bisphosphate. *Biochemistry* **30**, 7105–11 (1991).

[66] Zhong, W. *et al.* Allosteric pyruvate kinase-based "logic gate" synergistically senses energy and sugar levels in mycobacterium tuberculosis. *Nat Commun* **8**, 1986 (2017).

[67] Yang, J. *et al.* Synergistic allosteric mechanism of fructose-1,6-bisphosphate and serine for pyruvate kinase m2 via dynamics fluctuation network analysis. *J Chem Inf Model* **56**, 1184–1192 (2016).

[68] Xie, J., Dai, C. & Hu, X. Evidence that does not support pyruvate kinase m2 (PKM2)-catalyzed reaction as a rate-limiting step in cancer cell glycolysis. *J Biol Chem* **291**, 8987–99 (2016).

[69] Shestov, A. A. *et al.* Quantitative determinants of aerobic glycolysis identify flux through the enzyme gapdh as a limiting step. *Elife* **3** (2014).

[70] Hofmann, E. *et al.* Association-dissociation behavior of erythrocyte phosphofructokinase and tumor pyruvate kinase. *Adv Enzyme Regul* **13**, 247–77 (1975).

[71] Ashizawa, K., Kato, H., McPhie, P. & Cheng, S. Regulation of thyroid hormone binding to its cytosolic binding protein by L-alpha-alanine. *Biochem Biophys Res Commun* **167**, 587–92 (1990).

[72] Morgan, H. P. *et al.* Allosteric mechanism of pyruvate kinase from leishmania mexicana uses a rock and lock model. *J Biol Chem* **285**, 12892–8 (2010).

[73] Morgan, H. P. *et al.* Structures of pyruvate kinases display evolutionarily divergent allosteric strategies. *R Soc Open Sci* **1**, 140120 (2014).

[74] Naithani, A., Taylor, P., Erman, B. & Walkinshaw, M. D. A molecular dynamics study of allosteric transitions in leishmania mexicana pyruvate kinase. *Biophys J* **109**, 1149–56 (2015).

[75] Donovan, K. A. *et al.* Conformational dynamics and allostery in pyruvate kinase. *J Biol Chem* **291**, 9244–56 (2016).

[76] Tang, Q. & Fenton, A. W. Whole-protein alanine-scanning mutagenesis of allostery: A

large percentage of a protein can contribute to mechanism. *Hum Mutat* **38**, 1132–1143 (2017).

[77] Gehrig, S. *et al.* An engineered photoswitchable mammalian pyruvate kinase. *FEBS J* **284**, 2955–2980 (2017).

[78] Wang, F. *et al.* SIRT5 esuccinylates and activates pyruvate kinase M2 to block macrophage IL-1β production and to prevent DSS-induced colitis in mice. *Cell Rep* **19**, 2331–2344 (2017).

[79] Wang, Y. *et al.* O-glcnacylation destabilizes the active tetrameric PKM2 to promote the warburg effect. *Proc Natl Acad Sci USA* **114**, 13732–13737 (2017).

[80] Qi, W. *et al.* Pyruvate kinase m2 activation may protect against the progression of diabetic glomerular pathology and mitochondrial dysfunction. *Nat Med* **23**, 753–762 (2017).

[81] Lim, S.-O. *et al.* EGFR signaling enhances aerobic glycolysis in triple-negative breast cancer cells to promote tumor growth and immune escape. *Cancer Res* **76**, 1284–96 (2016).

[82] Hitosugi, T. *et al.* Tyrosine phosphorylation inhibits PKM2 to promote the warburg effect and tumor growth. *Sci Signal* **2**, 40–73 (2009).

[83] Macpherson, J. A. & Anastasiou, D. Allosteric regulation of metabolism in cancer: endogenous mechanisms and considerations for drug design. *Curr Opin Biotechnol* **48**, 102–110 (2017).

[84] Keedy, D. A. *et al.* An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife* **7** (2018).

[85] Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining markov state models and experiments. *Proc Natl Acad Sci USA* **112**, 2734–9 (2015).

[86] Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery.

Nature **508**, 331–9 (2014).

[87] del Sol, A., Tsai, C.-J., Ma, B. & Nussinov, R. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* **17**, 1042–50 (2009).

[88] Alberding, N. *et al.* Tunneling in ligand binding to heme proteins. *Science* **192**, 1002–4 (1976).

[89] Austin, R. H. *et al.* Dynamics of carbon monoxide binding by heme proteins. *Science* **181**, 541–3 (1973).

[90] Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–73 (1975).

[91] KENDREW, J. C. *et al.* Structure of myoglobin: A three-dimensional fourier synthesis at 2 a. resolution. *Nature* **185**, 422–7 (1960).

[92] Frauenfelder, H. Principles of ligand binding to heme proteins. *Methods Enzymol* **54**, 506–32 (1978).

[93] Frauenfelder, H., Fenimore, P. W. & Young, R. D. Protein dynamics and function: insights from the energy landscape and solvent slaving. *IUBMB Life* **59**, 506–12 (2007).

[94] Artymiuk, P. J. *et al.* Crystallographic studies of the dynamic properties of lysozyme. *Nature* **280**, 563–8 (1979).

[95] Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558–63 (1979).

[96] Changeux, J. P. The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb Symp Quant Biol* **26**, 313–8 (1961).

[97] Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J Mol Biol* **6**, 306–29 (1963).

[98] Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88–118 (1965).

[99] Koshland, D. E., Jr, Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–85 (1966).

[100] Changeux, J.-P. 50 years of allosteric interactions: the twists and turns of the models. *Nat Rev Mol Cell Biol* **14**, 819–29 (2013).

[101] Cornish-Bowden, A. Understanding allosteric and cooperative interactions in enzymes. *FEBS J* **281**, 621–32 (2014).

[102] Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci* **17**, 1295–307 (2008).

[103] Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726–39 (1970).

[104] Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646–52 (2002).

[105] Gelin, B. R. & Karplus, M. Sidechain torsional potentials and motion of amino acids in porteins: bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* **72**, 2002–6 (1975).

[106] Gelin, B. R. & Karplus, M. Mechanism of tertiary structural change in hemoglobin. *Proc Natl Acad Sci USA* **74**, 801–5 (1977).

[107] Rahman, A. Correlations in the motion of atoms in liquid argon. *Physical Review* **136**, A405 (1964).

[108] Stillinger, F. H. & Rahman, A. Improved simulation of liquid water by molecular dynamics. *J Chem Phys* **60**, 1545–1557 (1974).

[109] McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–90 (1977).

[110] Cooper, A. & Dryden, D. T. Allostery without conformational change. a plausible model. *Eur Biophys J* **11**, 103–9 (1984).

[111] Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. Dynamically driven protein allostery. *Nat Struct Mol Biol* **13**, 831–8 (2006).

[112] Vendruscolo, M. & Dobson, C. M. Towards complete descriptions of the free-energy landscapes of proteins. *Philos Trans A Math Phys Eng Sci* **363**, 433–452 (2005).

[113] Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429–33 (2001).

[114] Yao, X., Rosen, M. K. & Gardner, K. H. Estimation of the available free energy in a LOV2-J alpha photoswitch. *Nat Chem Biol* **4**, 491–7 (2008).

[115] Bürgi, R., Pitera, J. & van Gunsteren, W. F. Assessing the effect of conformational averaging on the measured values of observables. *J Biomol NMR* **19**, 305–20 (2001).

[116] Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–32 (2005).

[117] Guerry, P. *et al.* Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. *Angew Chem Int Ed Engl* **52**, 3181–5 (2013).

[118] Hünenberger, P. H., Mark, A. E. & van Gunsteren, W. F. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* **252**, 492–503 (1995).

[119] Ichiye, T. & Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205–17 (1991).

[120] Lange, O. F. & Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins* **62**, 1053–61 (2006).

[121] McClendon, C. L., Kornev, A. P., Gilson, M. K. & Taylor, S. S. Dynamic architecture of a protein kinase. *Proc Natl Acad Sci USA* **111**, 4623–4631 (2014).

[122] Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R. L. & McCammon, J. A.

Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci USA* **109**, 21216–22 (2012).

[123] Pandini, A., Fornili, A., Fraternali, F. & Kleinjung, J. Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *FASEB J* **26**, 868–81 (2012).

[124] Pandini, A., Fornili, A., Fraternali, F. & Kleinjung, J. GSATools: analysis of allosteric communication and functional local motions using a structural alphabet. *Bioinformatics* **29**, 2053–5 (2013).

[125] Pandini, A., Fornili, A. & Kleinjung, J. Structural alphabets derived from attractors in conformational space. *BMC Bioinformatics* **11**, 97 (2010).

[126] Hashem, S., Tiberti, M. & Fornili, A. Allosteric modulation of cardiac myosin dynamics by omecamtiv mecarbil. *PLoS Comput Biol* **13**, 1005–1026 (2017).

[127] Fornili, A., Pandini, A., Lu, H.-C. & Fraternali, F. Specialized dynamical properties of promiscuous residues revealed by simulated conformational ensembles. *J Chem Theory Comput* **9**, 5127–5147 (2013).

[128] Motta, S., Minici, C., Corrada, D., Bonati, L. & Pandini, A. Ligand-induced perturbation of the HIF-2α:ARNT dimer dynamics. *PLoS Comput Biol* **14**, 1006–1021 (2018).

[129] Pandini, A., Kleinjung, J., Rasool, S. & Khan, S. Coevolved mutations reveal distinct architectures for two core proteins in the bacterial flagellar motor. *PLoS One* **10**, 14–24 (2015).

[130] Pandini, A., Morcos, F. & Khan, S. The gearbox of the bacterial flagellar motor switch. *Structure* **24**, 1209–20 (2016).

[131] Yang, J.-S., Seo, S. W., Jang, S., Jung, G. Y. & Kim, S. Rational engineering of enzyme allosteric regulation through sequence evolution analysis. *PLoS Comput Biol* **8**, 1002–1012 (2012).

[132] Karplus, M. Molecular dynamics of biological macromolecules: a brief history and perspective. *Biopolymers* **68**, 350–8 (2003).

[133] Martyna, G. J., Tuckerman, M. E., Tobias, D. J. & Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol Phys* **87**, 1117–1157 (1996).

[134] Frenkel, D. & Smit, B. Understanding molecular dynamics simulations. *Academic, New York* (2001).

[135] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* **81**, 511–519 (1984).

[136] Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Physical review A* **31**, 1695 (1985).

[137] Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177–4189 (1994).

[138] Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *J Comput Chem* **25**, 1656–76 (2004).

[139] Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089–10092 (1993).

[140] Van Der Spoel, D. *et al.* Gromacs: fast, flexible, and free. *J Comput Chem* **26**, 1701–18 (2005).

[141] Berendsen, H., Grigera, J. & Straatsma, T. The missing term in effective pair potentials. *J Phys Chem* **91**, 6269–6271 (1987).

[142] Fiser, A. & Šali, A. Modeller: generation and refinement of homology-based protein structure models. In *Methods in enzymology*, vol. 374, 461–491 (Elsevier, 2003).

[143] Malde, A. K. *et al.* An automated force field topology builder (ATB) and repository: Version 1.0. *J Chem Theory Comput* **7**, 4026–37 (2011).

[144] Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* **4**, 116–22 (2008).

[145] Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* **13**, 952–962 (1992).

[146] Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved peptide and protein torsional energetics with the oplsaa force field. *J Chem Theory Comput* **11**, 3499–509 (2015).

[147] Daura, X. *et al.* The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and md simulation. *J Am Chem Soc* **123**, 2393–404 (2001).

[148] Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* **3**, 1139–52 (2008).

[149] Shvartsburg, A. A., Mashkevich, S. V., Baker, E. S. & Smith, R. D. Optimization of algorithms for ion mobility calculations. *J Phys Chem A* **111**, 2002–10 (2007).

[150] Hollenberg, P. F., Flashner, M. & Coon, M. J. Role of lysyl epsilon-amino groups in adenosine diphosphate binding and catalytic activity of pyruvate kinase. *J Biol Chem* **246**, 946–53 (1971).

[151] Baici, A. *Kinetics of Enzyme-Modifier Interactions - Selected Topics in the Theory and Diagnosis of Inhibition and Activation Mechanisms* (Springer, 2015).

[152] Reinhart, G. D. Quantitative analysis and interpretation of allosteric behavior. *Methods Enzymol* **380**, 187–203 (2004).

[153] Pacholarz, K. J. *et al.* Dynamics of intact IgG explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew Chem Int Ed Engl* **53**, 7765–9 (2014).

[154] Koler, R. D. & Vanbellinghen, P. The mechanism of precursor modulation of human pyruvate kinase i by fructose diphosphate. *Adv Enzyme Regul* **6**, 127–42 (1968).

[155] Kuczenski, R. T. & Suelter, C. H. Interactions of fructose 1,6-diphosphate, substrates, and monovalent cations with yeast pyruvate kinase monitored by changes in enzyme fluorescence. *Biochemistry* **10**, 2862–6 (1971).

[156] Bergmeyer, H.-U. *Methods of Enzymatic Analysis, Methods of Enzymatic Analysis: Volume 8: Metabolites 3: Lipids, Amino Acids and Related Compounds*, vol. 8 (Wiley-Blackwell, 1985).

[157] Hulme, E. C. & Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *Br J Pharmacol* **161**, 1219–37 (2010).

[158] Pollard, T. D. A guide to simple and informative binding assays. *Mol Biol Cell* **21**, 4061–7 (2010).

[159] Allali-Hassani, A. *et al.* A survey of proteins encoded by non-synonymous single nucleotide polymorphisms reveals a significant fraction with altered stability and activity. *Biochem J* **424**, 15–26 (2009).

[160] Gao, X. *et al.* Reciprocal regulation of protein kinase and pyruvate kinase activities of pyruvate kinase m2 by growth signals. *J Biol Chem* **288**, 15971–9 (2013).

[161] Jurica, M. S. *et al.* The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure* **6**, 195–210 (1998).

[162] Moreau, M. J., McGeoch, A. T., Lowe, A. R., Itzhaki, L. S. & Bell, S. D. Atpase site architecture and helicase mechanism of an archaeal mcm. *Mol Cell* **28**, 304–14 (2007).

[163] Rzechorzek, N. J. *et al.* Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. *Nat Commun* **5**, 5506 (2014).

[164] Barry, E. R., Lovett, J. E., Costa, A., Lea, S. M. & Bell, S. D. Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proc Natl Acad Sci USA* **106**, 1051–6 (2009).

[165] Crampton, D. J., Mukherjee, S. & Richardson, C. C. DNA-induced switch from independent to sequential dttp hydrolysis in the bacteriophage t7 DNA helicase. *Mol Cell*

21, 165–74 (2006).

[166] Jurneczko, E. & Barran, P. E. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **136**, 20–8 (2011).

[167] Pacholarz, K. J. *et al.* Hybrid mass spectrometry approaches to determine how L-histidine feedback regulates the enzyzme MtATP-Phosphoribosyltransferase. *Structure* **25**, 730–738.e4 (2017).

[168] Felíu, J. E. & Sols, A. Interconversion phenomena between two kinetic forms of class a pyruvate kinase from ehrlich ascites tumor cells. *Mol Cell Biochem* **13**, 31–44 (1976).

[169] Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol* **9**, 71 (2011).

[170] Guo, J. & Zhou, H.-X. Protein allostery and conformational dynamics. *Chem Rev* **116**, 6503–15 (2016).

[171] Winger, M., Trzesniak, D., Baron, R. & van Gunsteren, W. F. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Phys Chem Chem Phys* **11**, 1934–41 (2009).

[172] Fornili, A., Autore, F., Chakroun, N., Martinez, P. & Fraternali, F. Protein-water interactions in md simulations: Pops/popscomp solvent accessibility analysis, solvation forces and hydration sites. *Methods Mol Biol* **819**, 375–92 (2012).

[173] Capdevila, D. A., Braymer, J. J., Edmonds, K. A., Wu, H. & Giedroc, D. P. Entropy redistribution controls allostery in a metalloregulatory protein. *Proc Natl Acad Sci USA* **114**, 4424–4429 (2017).

[174] Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324–328 (2018).

[175] Tzeng, S.-R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488**, 236–40 (2012).

[176] Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem Phys Lett* **215**, 617–621 (1993).

[177] Craveur, P. *et al.* Protein flexibility in the light of structural alphabets. *Front Mol Biosci* **2**, 20 (2015).

[178] Wang, Z. & Moult, J. Snps, protein structure, and disease. *Hum Mutat* **17**, 263–70 (2001).

[179] Rosenzweig, R. & Kay, L. E. Bringing dynamic molecular machines into focus by methyl-trosy NMR. *Annu Rev Biochem* **83**, 291–315 (2014).

[180] Salvi, N., Abyzov, A. & Blackledge, M. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *J Phys Chem Lett* **7**, 2483–9 (2016).

[181] Markwick, P. R. L. *et al.* Toward a unified representation of protein structural dynamics in solution. *J Am Chem Soc* **131**, 16968–75 (2009).

[182] Schmidt, C. *et al.* Surface accessibility and dynamics of macromolecular assemblies probed by covalent labeling mass spectrometry and integrative modeling. *Anal Chem* **89**, 1459–1468 (2017).

[183] Kern, D. & Zuiderweg, E. R. P. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* **13**, 748–57 (2003).

[184] Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–21 (2005).

[185] Salvi, N., Papadopoulos, E., Blackledge, M. & Wagner, G. The role of dynamics and allostery in the inhibition of the eif4e/eif4g translation initiation factor complex. *Angew Chem Int Ed Engl* **55**, 7176–9 (2016).

[186] Micheletti, C., Seno, F. & Maritan, A. Recurrent oligomers in proteins: an optimal scheme reconciling accurate and concise backbone representations in automated folding and design studies. *Proteins* **40**, 662–74 (2000).

[187] Camproux, A. C., Gautier, R. & Tufféry, P. A hidden markov model derived structural alphabet for proteins. *J Mol Biol* **339**, 591–605 (2004).

[188] Ankerst, M., Breunig, M. M., Kriegel, H.-P. & Sander, J. Optics: ordering points to identify the clustering structure. In *ACM Sigmod record*, vol. 28, 49–60 (ACM, 1999).

[189] Freiburger, L. *et al.* Substrate-dependent switching of the allosteric binding mechanism of a dimeric enzyme. *Nat Chem Biol* **10**, 937–42 (2014).

[190] Hewitt, W. M. *et al.* Insights into the allosteric inhibition of the sumo e2 enzyme ubc9. *Angew Chem Int Ed Engl* **55**, 5703–7 (2016).

[191] Webb, B. A. *et al.* Structures of human phosphofructokinase-1 and atomic basis of cancer-associated mutations. *Nature* **523**, 111–4 (2015).

[192] Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–9 (1999).

[193] Teşileanu, T., Colwell, L. J. & Leibler, S. Protein sectors: statistical coupling analysis versus conservation. *PLoS Comput Biol* **11**, 1004–1019 (2015).

[194] Guarnera, E. & Berezovsky, I. N. Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS Comput Biol* **12**, 1004–1028 (2016).

[195] Guarnera, E., Tan, Z. W., Zheng, Z. & Berezovsky, I. N. Allosigma: allosteric signaling and mutation analysis server. *Bioinformatics* **33**, 3996–3998 (2017).

[196] Tee, W.-V., Guarnera, E. & Berezovsky, I. N. Reversing allosteric communication: From detecting allosteric sites to inducing and tuning targeted allosteric response. *PLoS Comput Biol* **14**, 1006–1028 (2018).

[197] de Juan, D., Pazos, F. & Valencia, A. Emerging methods in protein co-evolution. *Nat Rev Genet* **14**, 249–61 (2013).

[198] McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–42 (2012).

[199] Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–75 (2011).